
UN RECORRIDO POR LAS DISTRIBUCIONES DE PROBABILIDAD

María Laura Nores

RESUMEN. Las variables aleatorias tienen una distribución de probabilidad que bajo ciertas circunstancias viene dada en términos de una expresión conocida. Muchas veces esta fórmula se deduce fácilmente del contexto en el que surge la variable aleatoria, mientras que en otras ocasiones no resulta tan claro. En este artículo, se busca dar un paso más en la comprensión de las distribuciones de probabilidad más conocidas, intentando profundizar en la motivación que las origina y en la deducción de su expresión, destacando además relaciones existentes entre algunas distribuciones. De esta manera, el trabajo contribuye a brindar elementos para discernir cuándo resulta razonable asumir cierta distribución para una variable aleatoria en consideración.

ABSTRACT. Random variables have a probability distribution that under certain circumstances is given in terms of a known expression. Many times this formula is easily deduced from the context in which the random variable arises, while at other times it is not so clear. In this article, we seek to take a further step in understanding the most well-known probability distributions, trying to get deep into the motivation that gives them origin and in the deduction of their expression, also highlighting existing relationships between some distributions. In this way, the work contributes to provide elements to discern when it is reasonable to assume a certain distribution for a random variable under consideration.

§1. Introducción: Probabilidad y variables aleatorias

Los fenómenos aleatorios corresponden a situaciones en las que no hay certeza sobre cuál de los posibles resultados ocurrirá. Uno de los ejemplos más claros es tirar un dado: sabemos que saldrá algún valor entre 1 y 6, pero no podemos conocer de antemano cuál es el valor concreto que ocurrirá. Podemos pensar en otros ejemplos, como cuál es la altura o el peso de una persona elegida al azar, qué temperatura hará mañana o si un paciente mejorará o no al tomar cierto

Palabras clave: variable aleatoria discreta, variable aleatoria continua, densidad, independencia.
Keywords: discrete random variable, continuous random variable, density, independence.

medicamento. La Probabilidad se dedica al estudio de estos fenómenos aleatorios, intentando cuantificar la incertidumbre.

En principio parecería que no podemos hacer afirmaciones sobre fenómenos aleatorios. Volviendo al ejemplo del dado, no podemos predecir cuál será el valor de una o de unas pocas observaciones. Sin embargo, a medida que tenemos más observaciones, la proporción de veces (frecuencia relativa) que ocurre un resultado particular se va acercando cada vez más a un cierto número que sí esperaríamos: es razonable pensar que el número 6 saldrá $\frac{1}{6}$ de las veces si el número de tiradas del dado es lo suficientemente grande. Así, podemos pensar a la probabilidad de un resultado como la proporción de veces que ese resultado ocurriría si repitiéramos el experimento muchas veces.

Esta interpretación frecuentista de la probabilidad no es aplicable en todas las situaciones, ya que no siempre es posible repetir el experimento. Pensemos en una afirmación del tipo “el sospechoso 1 tiene una probabilidad de 0,8 de haber cometido el crimen contra el Sr. López”. Entonces, la probabilidad puede entenderse también como una medida del “grado de creencia” respecto a que un resultado pueda ocurrir, basado en la información disponible. A esta se la conoce como una visión “personal” o “subjetiva” de la probabilidad, y es la base de la Estadística Bayesiana.

La definición más moderna de la probabilidad se basa en una serie de axiomas, que intentan reflejar las propiedades que esta debería cumplir y que concuerdan con nuestra noción intuitiva de probabilidad, más allá de la interpretación que querramos darle (Ross, 2014). Veamos en qué consiste esta definición. Dado un fenómeno aleatorio, se llama espacio muestral al conjunto de todos los posibles resultados y se denota con Ω . Nosotros querríamos asignar probabilidad a subconjuntos del espacio muestral, denominados eventos, los cuales corresponden a un resultado particular (punto muestral) o a un grupo de posibles resultados. Así, dado un evento A en un espacio muestral Ω , llamaremos $P(A)$ a la probabilidad de que ocurra A , donde la función P satisface los siguientes axiomas:

- (i) $0 \leq P(A)$ para todo evento A .
- (ii) $P(\Omega) = 1$.
- (iii) Si A_1, A_2, \dots es una colección de eventos mutuamente excluyentes o disjuntos ($A_i \cap A_j = \emptyset$ si $i \neq j$), entonces

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Como consecuencia de estas condiciones, resulta que $P(\emptyset) = 0$, (iii) vale para una colección finita de eventos, $P(A^c) = 1 - P(A)$, con A^c el complemento de A , y $P(A) \leq 1$ para todo evento A . Notemos que estos axiomas no nos indican cómo asignar probabilidades a los eventos, solamente nos dicen cuáles son las

propiedades que debería cumplir esa asignación. Si en el ejemplo de tirar un dado asignamos probabilidad de $\frac{1}{6}$ a cada uno de los 6 resultados posibles y para cualquier evento A calculamos $P(A)$ como la suma de las probabilidades de los resultados incluidos en A , cumpliremos con las propiedades requeridas. Pero no *probamos* que cada resultado tenga probabilidad de $\frac{1}{6}$, partimos de una *asignación* de $\frac{1}{6}$ que hicimos en base a nuestra intuición y experiencia (Ash, 2008), que corresponde a pensar que todos los resultados son igualmente probables.

En un espacio muestral con un número finito de puntos, si todos los puntos tienen la misma probabilidad de ocurrencia, la probabilidad de un evento A se calcula como el número de casos favorables (número de puntos en el evento A) sobre número de casos posibles (número de puntos en el espacio muestral). Pero en el caso general esto no es aplicable.

Cuando hablamos de probabilidad, un concepto que no podemos dejar de mencionar es el de *probabilidad condicional*. ¿Qué información adicional nos da saber que ocurrió un evento A ? ¿Cómo afecta esto a la probabilidad de ocurrencia de otro evento, digamos, B ? ¿Siguiendo siendo la misma, es mayor o es menor? Si $P(A) > 0$, definimos la probabilidad condicional de B dado A como

$$P(B | A) = \frac{P(B \cap A)}{P(A)}.$$

La interpretación de la probabilidad condicional en términos de frecuencias relativas es intuitiva: si repetimos el experimento muchas veces, calculamos la proporción de veces que ocurrió B restringiéndonos a los casos en que ocurrió A . Esto es, número de veces que ocurrió A y B sobre número de veces que ocurrió A , que también puede verse como proporción de veces que ocurrió $B \cap A$ sobre proporción de veces que ocurrió A . Estas proporciones, como vimos recién, tienden a $P(B \cap A)$ y $P(A)$, respectivamente. Así, si sabemos que al arrojar un dado el resultado fue par, la probabilidad condicional de que sea 6 es $\frac{1/6}{3/6} = \frac{1}{3}$.

Notemos que si la ocurrencia de A no tiene efecto en la ocurrencia de B , esto es $P(B | A) = P(B)$, resulta por definición que $P(B \cap A) = P(B)P(A)$ y la recíproca también es válida. Así, diremos que dos eventos son *independientes* si la probabilidad de la intersección es el producto de las probabilidades. Esto se extiende a n eventos: diremos que los eventos A_1, \dots, A_n son independientes si la probabilidad de la intersección de cualquier subconjunto de estos eventos es el producto de las probabilidades de los eventos involucrados.

Luego de este repaso de las nociones de probabilidad, vamos a introducir el concepto de *variable aleatoria*. En un fenómeno aleatorio, muchas veces no estamos interesados en los resultados en sí mismos sino en una cantidad numérica determinada por el resultado, una función del resultado. Por ejemplo, si tiramos dos dados, quizás no nos interese el valor que toma cada dado sino la suma de ambos

valores. Si elegimos una persona al azar de una población, nos puede interesar su altura, su peso, si tiene o no obra social, etc.

Así, una variable aleatoria es una medición numérica del resultado de un fenómeno aleatorio. Formalmente, es una función

$$Y : \Omega \rightarrow \mathbb{R}$$

que asigna un valor numérico a cada punto del espacio muestral, a cada posible resultado del experimento. En general, se utilizan letras mayúscula de imprenta para denotar a las variables aleatorias (X, Y, Z, \dots) y minúscula para los valores que pueden tomar.

Dado que una variable aleatoria se refiere al resultado de un fenómeno aleatorio, cada valor posible que asume la variable aleatoria tiene una probabilidad específica de ocurrencia. La distribución de probabilidad de una variable aleatoria especifica sus posibles valores y sus correspondientes probabilidades ([Agresti, Franklin, y Klingenberg, 2018](#)).

Para $y \in \mathbb{R}$, se define el evento $\{Y = y\}$ como

$$\{Y = y\} = \{\omega \in \Omega : Y(\omega) = y\},$$

es decir, el conjunto de puntos del espacio muestral que van a parar al valor y por medio de la variable aleatoria Y . Luego, la probabilidad de que Y tome el valor y , $P(Y = y)$, corresponde a la probabilidad de que ocurra ese evento. De manera similar, $P(Y \leq y)$ es la probabilidad del evento formado por los puntos del espacio muestral que, al aplicarles la función Y , toman un valor menor o igual a y . La función F que a cada valor y en \mathbb{R} le asigna esta probabilidad, esto es,

$$F(y) = P(Y \leq y),$$

se denomina función de distribución (o función de distribución acumulada) de Y . Se puede probar que F cumple las siguientes propiedades (ver, por ejemplo, [Hoel, Port, y Stone \(1971\)](#)):

- $0 \leq F(y) \leq 1$ para todo y en \mathbb{R} .
- F es no decreciente (si $x < y \Rightarrow F(x) \leq F(y)$).
- $\lim_{y \rightarrow -\infty} F(y) = 0$.
- $\lim_{y \rightarrow \infty} F(y) = 1$.
- F es continua por derecha ($\lim_{y \rightarrow y_0^+} F(y) = F(y_0)$ para todo y_0 en \mathbb{R}).

Cuando se quiere distinguir la función de distribución de una variable aleatoria Y respecto de la de otras variables aleatorias, se suele usar la notación F_Y .

Una de las medidas de tendencia central más utilizadas para describir una variable aleatoria es, sin dudas, la *media*. Desde una interpretación frecuentista,

es el valor que obtendríamos para el promedio de los valores observados de esa variable aleatoria, si repitiéramos el experimento muchas veces. Pensemos en una variable Y que toma un número finito de valores y_1, \dots, y_r con probabilidades p_1, \dots, p_r , respectivamente, y que el experimento se realiza N veces. Luego, el promedio de los valores observados de Y sería

$$\frac{y_1 N_1 + \dots + y_r N_r}{N},$$

donde N_i es la cantidad de veces que ocurre el resultado y_i y $\sum_{i=1}^r N_i = N$. Como $\frac{N_i}{N}$ tiende a p_i , pues es la proporción de veces que ocurre el resultado y_i , el promedio se acerca al valor $y_1 p_1 + \dots + y_r p_r$, que es como se define la media de Y . Así, la media es un promedio ponderado de los valores que asume la variable, donde las ponderaciones son las probabilidades de ocurrencia de esos valores, de manera que los valores que son más probables reciben mayor peso. Es importante recalcar que la media no refleja el valor que obtendríamos en una observación individual, sino lo que esperaríamos para el promedio en una larga cantidad de observaciones (Agresti y cols., 2018). La media de una variable aleatoria Y también recibe el nombre de valor medio, valor esperado o esperanza de Y y se denota $E(Y)$ o con la letra griega μ .

Además de dar una medida de centralidad como es la media, es conveniente conocer el grado de dispersión de la variable. Para cuantificar la variabilidad de Y , podemos pensar en la variable aleatoria que mide las desviaciones respecto a la media, esto es, $Y - E(Y)$. Como estas desviaciones pueden ser positivas o negativas, para evitar el signo elevamos esta variable al cuadrado y luego calculamos su media. Así, obtenemos una medida de dispersión o variabilidad conocida como *varianza*:

$$Var(Y) = E((Y - E(Y))^2).$$

Usualmente se denota con σ^2 . A mayor concentración de los valores de Y en torno a la media μ , menor σ^2 . Tomando la raíz cuadrada de la varianza, obtenemos otra medida de variabilidad conocida como *desvío* o *desviación estándar*, denotada con σ .

Existen distintos tipos de variables aleatorias según los valores que pueden asumir. En las siguientes secciones veremos definiciones un poco más precisas, pero por ahora daremos una idea de la clasificación de variables aleatorias en discretas y continuas.

Una variable aleatoria se dice *discreta* si toma un conjunto de valores que están “separados” entre sí. En muchos casos están asociadas a conteos y los valores posibles son los números naturales o un subconjunto de ellos. Por ejemplo, cantidad de hijos, número de accidentes en una ruta en un año, número de pacientes que responden a una medicación de un total de pacientes que la recibieron, etc. Por otro lado, una variable aleatoria es *continua* si puede tomar valores que forman un intervalo en vez de un conjunto de valores puntuales. Por ejemplo, la altura, el peso,

la distancia recorrida hasta lograr un determinado objetivo, el tiempo en realizar una tarea, etc. Veremos que la distribución de una variable aleatoria continua se especifica por medio de una curva que permite determinar la probabilidad de que la variable aleatoria caiga en un intervalo particular de valores, la cual está dada por el área bajo la curva en ese intervalo.

Si bien la distribución de una variable aleatoria en principio podría estar dada en términos de cualquier expresión, existen distribuciones que han sido ampliamente estudiadas porque se aplican en muchas situaciones con características comunes. Para algunas distribuciones de probabilidad conocidas, en la mayoría de los cursos de Probabilidad se presenta una motivación y a partir de la misma se deduce una expresión para el cálculo de las probabilidades asociadas a la variable aleatoria. Este suele ser el caso de las distribuciones binomial, geométrica, binomial negativa e hipergeométrica. Pero para otras distribuciones como la de Poisson, exponencial, Gamma, se suele presentar una fórmula y solamente se comenta en qué situaciones puede ser conveniente asumir una distribución de ese tipo, sin una justificación clara. En este artículo estudiaremos distribuciones que son familiares para el lector que haya tenido un curso de Probabilidad, con el objetivo de motivar y comprender más a fondo su definición y volver más tangible su uso. También, con el mismo fin, se presentan algunos resultados interesantes que conectan a algunas distribuciones entre sí.

A continuación, emprendemos nuestro recorrido por las distintas distribuciones de probabilidad, comenzando por las variables aleatorias discretas para luego estudiar las continuas.

§2. Variables aleatorias discretas

Formalmente, una variable aleatoria Y se dice discreta si asume una cantidad finita de valores o infinita numerable (es decir, los valores que puede tomar se pueden poner en correspondencia con los números naturales). Se llama función de densidad discreta, función de probabilidad (puntual) o distribución de probabilidad de Y a la función $f : \mathbb{R} \rightarrow \mathbb{R}$ definida como

$$f(y) = P(Y = y) \quad \text{para todo } y \text{ en } \mathbb{R}.$$

También se la llama función de frecuencia o función de masa de probabilidad y se suele usar la notación $p(y)$. Así, la función asigna a cada valor real la probabilidad de que la variable Y asuma ese valor.

Como el conjunto de valores que puede tomar Y (rango de Y) es numerable, podemos escribirlo como $\mathcal{R}(Y) = \{y_1, y_2, \dots\}$. Luego, f toma valor 0 para y fuera de ese conjunto, ya que $\{Y = y\} = \emptyset$ en ese caso. Además, $\sum_i f(y_i) = 1$, ya que la unión de todos los eventos $\{Y = y_i\}$ forma el espacio muestral Ω y P cumple los axiomas (ii) y (iii) por ser una función de probabilidad. A los valores que toma la

variable Y con probabilidad diferente de 0 se los suele llamar valores posibles de Y .

El valor esperado de Y se define como

$$E(Y) = \sum_i y_i f(y_i)$$

si $\sum_i |y_i| f(y_i) < \infty$, en cuyo caso decimos que Y tiene esperanza finita.

Si Y^2 tiene esperanza finita, se define la varianza de Y como

$$(2.1) \quad \text{Var}(Y) = E(Y - E(Y))^2.$$

Utilizando propiedades de la esperanza, se puede deducir la siguiente fórmula útil para el cálculo de la varianza:

$$\text{Var}(Y) = E(Y^2) - E(Y)^2,$$

donde $E(Y^2)$ puede calcularse como $E(Y^2) = \sum_i y_i^2 f(y_i)$.

La función de distribución F tiene un aspecto muy característico en las variables aleatorias discretas: si $y_1 < y_2 < y_3 < \dots$, F tiene forma de escalera, constante en $(y_i, y_{i+1}]$, con un salto en y_i de magnitud $f(y_i) = P(Y = y_i)$. Un ejemplo de esto puede verse en la Figura 1 B), correspondiente a la distribución binomial, que veremos a continuación.

2.1. La distribución binomial. Empezaremos nuestro recorrido por la distribución binomial, una de las más conocidas para variables aleatorias discretas. Surge en el contexto de un experimento binomial, el cual está definido por lo siguiente (Devore, 2005; Wackerly, Mendenhall, y Scheaffer, 2010):

- El experimento consiste en una secuencia de un número fijo n de ensayos independientes (repeticiones de un experimento más pequeño, donde el resultado de cada uno de estos experimentos no tiene influencia en el resultado de los otros).
- Cada ensayo puede dar como resultado uno de dos resultados posibles, generalmente denotados como éxito y fracaso (ejemplos: sí/no, presencia/ausencia, 1/0).
- La probabilidad de éxito es la misma en todos los ensayos y se denota con p .

Uno de los ejemplos más claros consiste en tirar n veces una moneda: en cada tirada se puede obtener cara (éxito) o cruz (fracaso) y el resultado obtenido en cada tirada no influye en el resto de las tiradas. Además, la probabilidad de obtener cara en una tirada cualquiera es siempre p (con $p = 0,5$ si la moneda es "honest").

Cabe mencionar que la denominación "éxito" no tiene que tener necesariamente una connotación positiva; puede referirse a cualquiera de los dos resultados posibles, pero hay que especificar a cuál de ellos ya que p corresponde a la probabilidad de éxito (y por lo tanto $1 - p$, a la probabilidad de fracaso).

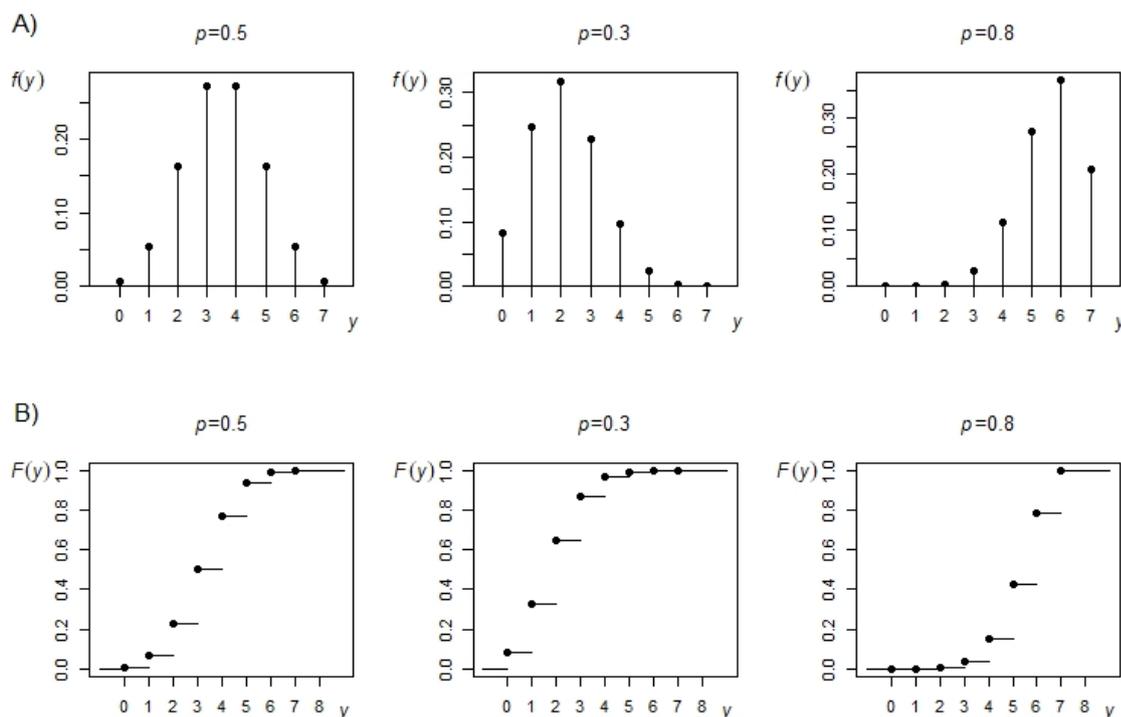


FIGURA 1. Gráficas de la función de masa de probabilidad (A) y la función de distribución (B) de una variable aleatoria con distribución $B(7, p)$ para distintos valores de p .

Para un experimento binomial, un punto ω en el espacio muestral puede representarse por una n -upla donde en la posición i hay un 1 si el ensayo i es un éxito y hay un 0 si es fracaso.

Consideremos ahora la variable aleatoria Y que cuenta el número de éxitos entre los n ensayos independientes. ¿Cuál es la distribución de Y ? Claramente Y toma valores en el conjunto $\{0, 1, \dots, n\}$. Tomemos k un número en este conjunto; estamos interesados en calcular $P(Y = k)$. Notemos que un punto ω está en el evento $\{Y = k\}$ si tiene k unos y $n - k$ ceros. Imaginemos, sin pérdida de generalidad, que los unos están en los primeros k lugares. Esto es,

$$\omega = (\underbrace{1, \dots, 1}_k, \underbrace{0, \dots, 0}_{n-k}).$$

Si definimos, para $i = 1, \dots, n$, el evento

$$A_i = \{\text{el ensayo } i \text{ es éxito}\},$$

ω corresponde al evento

$$A_1 \cap A_2 \cap \dots \cap A_k \cap A_{k+1}^c \cap A_{k+2}^c \cap \dots \cap A_n^c,$$

ya que simultáneamente ocurren los n eventos. Además, como estos eventos son independientes, resulta

$$P(\{\omega\}) = P(A_1)P(A_2) \cdots P(A_k)P(A_{k+1}^c)P(A_{k+2}^c) \cdots P(A_n^c).$$

Como en cada ensayo la probabilidad de éxito es p y la probabilidad de fracaso es $1 - p$, resulta que

$$P(\{\omega\}) = p^k(1 - p)^{n-k},$$

y esto es así para todo ω en el evento $\{Y = k\}$. Ahora, ¿cuántos puntos hay en este evento? Tantos como formas hay de elegir k lugares (para poder ubicar los unos) entre n , esto es, el número combinatorio $\binom{n}{k}$. Así,

$$P(Y = k) = \binom{n}{k}p^k(1 - p)^{n-k}.$$

Luego, la función de densidad discreta de Y es

$$f(y) = \begin{cases} \binom{n}{y}p^y(1 - p)^{n-y}, & y = 0, 1, \dots, n, \\ 0, & \text{caso contrario.} \end{cases}$$

Se dice que Y tiene distribución binomial con parámetros n y p y se denota por

$$Y \sim B(n, p).$$

Cuando $n = 1$, se dice que Y es una variable aleatoria de Bernoulli y toma valores 1 y 0 con probabilidades p y $1 - p$ respectivamente.

Ejemplo: Si tiramos 7 veces una moneda honesta, el número de caras en las 7 tiradas tiene distribución $B(7, \frac{1}{2})$. Si la moneda no es honesta y la probabilidad de obtener cara es un valor p entre 0 y 1, diferente de $\frac{1}{2}$, entonces el número de caras en las 7 tiradas tendrá distribución $B(7, p)$. La Figura 1 muestra la gráfica de la función de masa de probabilidad (A) y de distribución acumulada (B) de una distribución $B(7, p)$ para distintos valores de p .

Para ejemplificar el cálculo de esperanza y varianza de una variable aleatoria discreta, haremos el desarrollo para la distribución binomial.

Proposición: Sea $Y \sim B(n, p)$. Entonces, $E(Y) = np$ y $Var(Y) = np(1 - p)$.

Demostración: Para calcular $E(Y)$, notemos primero que $y\binom{n}{y} = 0$ para $y = 0$ y para $y = 1, \dots, n$ vale

$$y\frac{n!}{y(y-1)!(n-y)!} = \frac{n(n-1)!}{(y-1)!(n-1-(y-1))!} = n\binom{n-1}{y-1}.$$

Luego,

$$E(Y) = \sum_{y=0}^n y\binom{n}{y}p^y(1-p)^{n-y} = \sum_{y=1}^n n\binom{n-1}{y-1}p^y(1-p)^{n-y}$$

y por lo tanto

$$\begin{aligned} E(Y) &= np \sum_{y=1}^n \binom{n-1}{y-1} p^{y-1} (1-p)^{n-1-(y-1)} \\ &= np \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} = np(p + (1-p))^{n-1} = np, \end{aligned}$$

tomando $i = y - 1$. Otra forma de ver que la última sumatoria da 1, es notar que estamos sumando las probabilidades asociadas a una variable binomial con parámetros $n - 1$ y p . Así, el número esperado de éxitos entre n ensayos independientes es np , donde p es la probabilidad de éxito en cada ensayo individual.

Para calcular $E(Y^2)$, utilizaremos el siguiente truco que también sirve para otras distribuciones: escribimos y^2 como

$$y(y - 1 + 1) = y(y - 1) + y.$$

Como el término asociado a $y = 0$ es 0, resulta

$$\begin{aligned} E(Y^2) &= \sum_{y=0}^n y^2 \binom{n}{y} p^y (1-p)^{n-y} \\ &= \sum_{y=1}^n y(y-1) \binom{n}{y} p^y (1-p)^{n-y} + \sum_{y=1}^n y \binom{n}{y} p^y (1-p)^{n-y} \end{aligned}$$

y el segundo término es $E(Y) = np$. Ahora, $y(y-1) \binom{n}{y} = 0$ para $y = 1$, y para $y = 2, \dots, n$ vale

$$y(y-1) \frac{n(n-1)(n-2)!}{y(y-1)(y-2)!(n-2-(y-2))!} = n(n-1) \binom{n-2}{y-2}.$$

Luego,

$$\begin{aligned} E(Y^2) &= \sum_{y=2}^n n(n-1) \binom{n-2}{y-2} p^y (1-p)^{n-y} + np \\ &= n(n-1)p^2 \sum_{y=2}^n \binom{n-2}{y-2} p^{y-2} (1-p)^{n-2-(y-2)} + np \\ &= n(n-1)p^2 \sum_{i=0}^{n-2} \binom{n-2}{i} p^i (1-p)^{n-2-i} + np \\ &= n(n-1)p^2(p + (1-p))^{n-2} + np = n(n-1)p^2 + np, \end{aligned}$$

tomando $i = y - 2$. Aquí también podemos observar que la última sumatoria da 1 por ser suma de las probabilidades asociadas a una variable binomial con parámetros $n - 2$ y p . Finalmente,

$$\text{Var}(Y) = n(n-1)p^2 + np - (np)^2 = -np^2 + np = np(1-p),$$

como queríamos demostrar. \square

2.2. La distribución hipergeométrica. Sigamos nuestro recorrido por la distribución hipergeométrica, comenzando por entender en qué situación se origina. Consideremos una población con N objetos, de los cuales M son de tipo I y $N - M$ de tipo II. Se toma al azar una muestra de tamaño n (con $n \leq N$) *sin reposición*. ¿Qué distribución tiene la variable Y que cuenta el número de objetos de tipo I en la muestra? Como la muestra es elegida al azar, todos los posibles subconjuntos de n objetos tendrán la misma probabilidad $\frac{1}{\binom{N}{n}}$ de ser seleccionados. La cantidad de formas de elegir n objetos de los cuales y sean de tipo I y los restantes $n - y$ sean de tipo II es $\binom{M}{y} \binom{N-M}{n-y}$, con lo cual

$$P(Y = y) = \frac{\binom{M}{y} \binom{N-M}{n-y}}{\binom{N}{n}}.$$

Notar que esto tiene sentido en principio para y un entero entre 0 y n , ya que se eligen n objetos en total. Además, debe ser $y \leq M$ (porque no puede haber más objetos de tipo I en la muestra que en toda la población) y también $n - y \leq N - M$ (porque no puede haber más objetos de tipo II en la muestra que en toda la población), con lo cual debe ocurrir $y \geq n - N + M$. Así, se dice que una variable Y tiene distribución hipergeométrica de parámetros n , M y N y se denota

$$Y \sim \mathcal{H}(n, M, N)$$

si la función de densidad discreta está dada por

$$f(y) = \begin{cases} \frac{\binom{M}{y} \binom{N-M}{n-y}}{\binom{N}{n}}, & \text{máx}\{0, n - N + M\} \leq y \leq \text{mín}\{n, M\}, y \in \mathbb{Z}, \\ 0, & \text{caso contrario.} \end{cases}$$

Ejemplo (Wackerly y cols., 2010): Un producto industrial se envía en lotes de 20 unidades. El proceso de control de calidad requiere que se muestreen 5 artículos de cada lote y el lote completo será rechazado si contiene más de un artículo defectuoso. Si un lote contiene 4 artículos defectuosos, ¿cuál es la probabilidad de que sea rechazado?

Solución: Sea $Y =$ Número de productos defectuosos en la muestra de 5 artículos. Luego, $Y \sim \mathcal{H}(5, 4, 20)$. Así,

$$\begin{aligned} P(\text{rechazar el lote}) &= P(Y \geq 2) \\ &= 1 - P(Y = 0) - P(Y = 1) \\ &= 1 - \frac{\binom{4}{0} \binom{16}{5}}{\binom{20}{5}} - \frac{\binom{4}{1} \binom{16}{4}}{\binom{20}{5}} \\ &= 1 - 0,2817 - 0,4696 = 0,2487. \end{aligned}$$

2.3. Conexión entre las distribuciones binomial e hipergeométrica. Consideremos la siguiente situación, que nos hará reflexionar sobre similitudes y diferencias entre las distribuciones binomial e hipergeométrica:

Ejemplo: Una máquina produce tornillos, de los cuales el 10% son defectuosos. Encuentre la probabilidad de que una caja de 3 tornillos contenga a lo sumo uno defectuoso.

Solución: ¿Qué distribución tiene la variable aleatoria que cuenta el número de tornillos defectuosos en la caja? ¿Es binomial? ¿Es hipergeométrica? Sería binomial si la muestra se extrajera *con reposición*; es decir, cada vez que saco un tornillo de la producción total, lo vuelvo a poner antes de extraer otro tornillo. Así, cada tornillo extraído sería, independientemente de los otros tornillos, defectuoso o no defectuoso con probabilidades 0,1 y 0,9 respectivamente (las proporciones de tornillos defectuosos y no defectuosos en la producción total de tornillos). Pero aquí el muestreo es *sin reposición*, por lo cual si el primer tornillo es defectuoso, las proporciones de tornillos defectuosos y no defectuosos en el conjunto de los tornillos restantes cambian. Nuestra variable es en esencia hipergeométrica, pero como no contamos con el tamaño N de la población de la cual la muestra de 3 tornillos fue extraída, no podemos realizar el cálculo de las probabilidades. Lo que sí es razonable considerar es que N es grande, ya que la producción de tornillos no va a ser de unos pocos tornillos... Para ilustrar, supongamos que $N = 10000$. Si el 10% de los tornillos son defectuosos, entonces hay 1000 tornillos defectuosos y 9000 no defectuosos. Luego, la variable Y que cuenta el número de tornillos defectuosos en la muestra de 3 tornillos tiene distribución $\mathcal{H}(3, 1000, 10000)$.

Consideremos los eventos $B_0 = \{\text{ninguno de los tres tornillos es defectuoso}\}$ y $B_1 = \{\text{exactamente uno de los tres tornillos es defectuoso}\}$. Luego, la probabilidad buscada es

$$P(Y \leq 1) = P(B_0 \cup B_1) = P(B_0) + P(B_1),$$

con

$$P(B_0) = P(Y = 0) = \frac{\binom{1000}{0} \binom{9000}{3}}{\binom{10000}{3}} = \frac{9000 \times 8999 \times 8998}{10000 \times 9999 \times 9998} \simeq 0,9^3$$

y

$$P(B_1) = P(Y = 1) = \frac{\binom{1000}{1} \binom{9000}{2}}{\binom{10000}{3}} = 3 \frac{1000 \times 9000 \times 8999}{10000 \times 9999 \times 9998} \simeq 3 \times 0,1 \times 0,9^2.$$

Así,

$$P(Y \leq 1) \simeq 0,9^3 + 3 \times 0,1 \times 0,9^2 = P(X \leq 1)$$

para $X \sim B(3; 0,10)$. Es decir, podemos utilizar la distribución binomial para *aproxi-*
mar la probabilidad deseada, sin necesidad de conocer el tamaño de la población, sino solamente la proporción de éxitos en la población.

Veamos formalmente un resultado que muestra que la distribución hipergeométrica puede aproximarse por la distribución binomial si el tamaño de la muestra es pequeño en relación al tamaño de la población (Blitzstein y Hwang, 2015; Ross, 2014).

Aproximación a la distribución hipergeométrica por la distribución binomial:

Supongamos que $Y \sim \mathcal{H}(n, M, N)$ y que el tamaño de la población N crece de manera tal que la proporción $\frac{M}{N}$ de objetos tipo I se mantiene constante, igual a p . Luego,

$$P(Y = y) \xrightarrow{N \rightarrow \infty} \binom{n}{y} p^y (1-p)^{n-y}.$$

Demostración:

$$P(Y = y) = \frac{\binom{M}{y} \binom{N-M}{n-y}}{\binom{N}{n}} = \frac{M!}{y!(M-y)!} \frac{(N-M)!}{(n-y)!(N-M-n+y)!} \frac{n!(N-n)!}{N!}$$

de donde

$$P(Y = y) = \binom{n}{y} \frac{M(M-1)\cdots(M-y+1)(N-M)(N-M-1)\cdots(N-M-n+y+1)}{N(N-1)\cdots(N-n+1)}.$$

Notar que el numerador tiene $y + (n-y) = n$ factores, al igual que el denominador. Dividiendo cada factor de numerador y denominador por N y reemplazando $\frac{M}{N}$ por p se obtiene

$$P(Y = y) = \binom{n}{y} \frac{p(p - \frac{1}{N}) \cdots (p - \frac{y-1}{N})(1-p)(1-p - \frac{1}{N}) \cdots (1-p - \frac{n-y-1}{N})}{(1 - \frac{1}{N})(1 - \frac{2}{N}) \cdots (1 - \frac{n-1}{N})}$$

de donde se deduce que

$$P(Y = y) \xrightarrow{N \rightarrow \infty} \binom{n}{y} p^y (1-p)^{n-y},$$

que es la probabilidad de que una variable aleatoria binomial de parámetros n y p tome el valor y . \square

Luego, para N suficientemente grande en relación al tamaño de la muestra n , podemos aproximar a la distribución $\mathcal{H}(n, M, N)$ por la $B(n, \frac{M}{N})$, donde $p = \frac{M}{N}$ es la proporción de objetos tipo I ("éxitos") en la población. La aproximación es buena si $\frac{n}{N} < 0,05$ y p no está demasiado cerca de 0 ni de 1 (Devore, 2005). Así, a medida que el tamaño de la población N total aumenta mucho en relación al tamaño de la muestra que se extrae, el muestreo con reemplazo (que se corresponde con la distribución binomial) y el muestreo sin reemplazo (distribución hipergeométrica) se vuelven esencialmente equivalentes. Algo importante que volvemos a recalcar es que la aproximación binomial no requiere conocer el tamaño de la población, sino solamente la proporción de objetos de cada tipo.

Tabla 1: $P(X = x)$ para $x = 0, \dots, 5$, cuando $X \sim \mathcal{H}(n, M, N)$ y la correspondiente aproximación mediante la distribución $B(n, \frac{M}{N})$.

Distribución	x					
	0	1	2	3	4	5
$\mathcal{H}(5, 10, 100)$	0,5838	0,3394	0,0702	0,0064	0,0003	0,0000
$B(5; 0,1)$	0,5905	0,3281	0,0729	0,0081	0,0005	0,0000
$\mathcal{H}(10, 40, 200)$	0,1013	0,2683	0,3098	0,2052	0,0863	0,0240
$B(10; 0,2)$	0,1074	0,2684	0,3020	0,2013	0,0881	0,0264
$\mathcal{H}(10, 90, 300)$	0,0264	0,1184	0,2348	0,2714	0,2025	0,1020
$B(10; 0,3)$	0,0282	0,1211	0,2335	0,2668	0,2001	0,1029
$\mathcal{H}(5, 160, 400)$	0,0765	0,2592	0,3478	0,2309	0,0758	0,0099
$B(5; 0,4)$	0,0778	0,2592	0,3456	0,2304	0,0768	0,0102
$\mathcal{H}(7, 250, 500)$	0,0075	0,0537	0,1637	0,2751	0,2751	0,1637
$B(7; 0,5)$	0,0078	0,0547	0,1641	0,2734	0,2734	0,1641

Así, esta aproximación nos permite tratar como binomial al número de individuos en una muestra extraída al azar de una población (sin reposición, relativamente pequeña) que poseen una cierta característica, siendo p la proporción de individuos de la población que la poseen y n el tamaño de la muestra extraída. Por ejemplo, si se toma una muestra al azar de personas de una ciudad, se puede utilizar la distribución binomial para calcular la probabilidad de que haya una cierta cantidad de personas en la muestra que estén a favor de un cierto candidato, que fumen, que usen una determinada tarjeta de crédito, o que posean una cierta característica física (color específico de cabello, ojos, etc.). Además, si p es desconocido, podremos estimarlo a partir de la proporción de éxitos observada en la muestra.

La Tabla 1 ilustra cuán buena es la aproximación para distintos valores de los parámetros. Los cálculos de las probabilidades se realizaron utilizando el software R, versión 3.6.1 (R Core Team, 2019), mediante las funciones `dhyper` y `dbinom`.

2.4. La distribución de Poisson. Esta distribución ya no tiene una interpretación tan directa y generalmente es presentada en términos de la función de densidad discreta o función de masa de probabilidad, dada por

$$(2.2) \quad f(y) = \begin{cases} \frac{e^{-\lambda} \lambda^y}{y!}, & y = 0, 1, \dots, \\ 0, & \text{caso contrario} \end{cases}$$

En este caso, se dice que la variable aleatoria Y tiene distribución de Poisson de parámetro $\lambda > 0$ y se denota

$$Y \sim \mathcal{P}(\lambda).$$

Notemos que el rango de Y ya no es finito sino que es infinito numerable. ¿Cómo podemos llegar a la expresión dada en (2.2)? A continuación veremos que la distribución de Poisson se puede utilizar como aproximación de la distribución binomial de parámetros n y p cuando n es, de manera relativa, grande, la probabilidad de éxito p es pequeña y $\lambda = np$ toma un valor moderado (Canavos, 1995). El resultado es el siguiente (Wackerly y cols., 2010; Ross, 2014; Hoel y cols., 1971).

Aproximación a la distribución binomial por la distribución de Poisson: Supongamos $Y \sim B(n, p)$ con $p = \frac{\lambda}{n}$, para algún $\lambda > 0$. Entonces,

$$P(Y = y) \xrightarrow{n \rightarrow \infty} \frac{e^{-\lambda} \lambda^y}{y!}, \quad y = 0, 1, \dots$$

Demostración:

$$\begin{aligned} P(Y = y) &= \binom{n}{y} p^y (1-p)^{n-y} = \frac{n(n-1) \cdots (n-y+1)}{y!} \frac{\lambda^y}{n^y} \left(1 - \frac{\lambda}{n}\right)^{n-y} \\ &= \frac{\lambda^y}{y!} \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{y-1}{n}\right) \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-y}. \end{aligned}$$

Como $\lim_{n \rightarrow \infty} \left(1 - \frac{i}{n}\right) = 1$ para $i = 1, \dots, y-1$, $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-y} = 1$ y $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$, resulta que $P(Y = y) \xrightarrow{n \rightarrow \infty} \frac{e^{-\lambda} \lambda^y}{y!}$, que es la probabilidad de que una variable aleatoria con distribución $\mathcal{P}(\lambda)$ tome el valor y . □

También puede obtenerse la aproximación de Poisson a la distribución binomial considerando que $n \rightarrow \infty$ y $p \rightarrow 0$ de manera tal que np tiende a un valor $\lambda > 0$ (Devore, 2005; Grimmett y Stirzaker, 2001). Como regla empírica, esta aproximación se puede aplicar cuando $n \geq 100$, $p \leq 0,01$ y $np \leq 20$ (Devore, 2005).

La Tabla 2 muestra la aproximación para distintos valores de n , p y $\lambda = np$. Las probabilidades se calcularon utilizando las funciones `dbinom` y `dpois` de R.

Tabla 2: $P(X = x)$ para $x = 0, \dots, 5$, cuando $X \sim B(n, p)$ y la correspondiente aproximación mediante la distribución de Poisson con $\lambda = np$.

Distribución	x					
	0	1	2	3	4	5
$B(100; 0,01)$	0,3660	0,3697	0,1849	0,0610	0,0149	0,0029
$\mathcal{P}(1)$	0,3679	0,3679	0,1839	0,0613	0,0153	0,0031
$B(200; 0,01)$	0,1340	0,2707	0,2720	0,1814	0,0902	0,0357
$\mathcal{P}(2)$	0,1353	0,2707	0,2707	0,1804	0,0902	0,0361
$B(500; 0,006)$	0,0493	0,1489	0,2243	0,2247	0,1685	0,1009
$\mathcal{P}(3)$	0,0498	0,1494	0,2240	0,2240	0,1680	0,1008
$B(1000; 0,004)$	0,0182	0,0730	0,1464	0,1956	0,1958	0,1566
$\mathcal{P}(4)$	0,0183	0,0733	0,1465	0,1954	0,1954	0,1563

La distribución de Poisson se utiliza frecuentemente para modelar la distribución de probabilidad del número de sucesos o eventos que ocurren en un determinado tiempo o espacio, como el número de entradas a una página web en un año, el número de personas que ingresan a un banco o a un comercio en un día, el número de árboles de una determinada especie por km^2 , entre otros. ¿Por qué esto es así? Se debe a que, como recién vimos, se puede aproximar a la distribución binomial por medio de la distribución de Poisson. Siguiendo a [Wackerly y cols. \(2010\)](#), pensemos en el número de eventos que ocurren en un intervalo de tiempo de longitud t , digamos $[0, t]$. Dividamos el intervalo $[0, t]$ en n subintervalos, con n suficientemente grande, de manera tal que la longitud de cada subintervalo sea tan pequeña que a lo sumo un evento podría ocurrir en él con probabilidad diferente de cero y consideremos que esta probabilidad es la misma para todos los subintervalos. Esto es:

- $P(\text{ocurre exactamente un evento en un subintervalo}) = p.$
- $P(\text{no ocurren eventos en un subintervalo}) = 1 - p.$
- $P(\text{ocurre más de un evento en un subintervalo}) = 0.$

Si la ocurrencia de eventos puede ser considerada como independiente de un subintervalo a otro, entonces el número de eventos que ocurren en $[0, t]$ coincide con el número de subintervalos en los que ocurre un evento y por lo tanto tiene distribución binomial con parámetros n y p . Si bien no hay una única forma de seleccionar los subintervalos y por lo tanto no conocemos ni n ni p , parece razonable pensar que cuando consideramos un número mayor de subintervalos (aumentamos el n), disminuye la probabilidad p de que ocurra un evento en cada subintervalo. Y por el resultado que vimos antes, surge naturalmente la distribución de Poisson para la variable aleatoria “Número de eventos que ocurren en un intervalo de longitud t ”. El parámetro λ en la distribución de Poisson es el valor esperado (media) de la variable aleatoria, por lo cual sería el número esperado de eventos, o valor promedio de eventos que ocurren en ese período de tiempo.

Con un razonamiento similar, podemos pensar que el número de eventos que ocurren en un determinado espacio (como el número de árboles por km^2) tiene distribución de Poisson. También podemos llegar a la expresión (2.2) relajando un poco los supuestos recién presentados, suponiendo que la probabilidad de que ocurra más de un evento en un subintervalo no es 0 sino que *tiende* a 0 ([Ross, 2014](#)). Este resultado, más realista pero más complejo a la vez, se presenta en el Apéndice.

2.5. La distribución geométrica. La distribución geométrica surge en el contexto de un experimento que comparte algunas de las características de un experimento binomial. Consiste en una secuencia de ensayos independientes, con dos resultados posibles: éxito o fracaso, donde la probabilidad de éxito es igual a p y es la misma

en todos los ensayos. Sin embargo, el número total de ensayos no está fijo sino que el experimento concluye con el primer éxito.

Consideremos la variable aleatoria Y que corresponde al número del ensayo en el que ocurre el primer éxito, o dicho de otra manera, al número total de ensayos hasta que ocurre el primer éxito. Claramente los valores posibles de Y son todos los números naturales, ya que el primer éxito podría ocurrir en el primer ensayo, en el segundo, en el tercero, etc. ¿Cuál será entonces la $P(Y = k)$, con $k \in \mathbb{N}$? Si definimos, como antes, el evento

$$A_i = \{\text{el ensayo } i \text{ es éxito}\},$$

entonces el evento $\{Y = k\}$ equivale a $A_1^c \cap A_2^c \cap \dots \cap A_{k-1}^c \cap A_k$, ya que que los primeros $k - 1$ ensayos necesariamente deben ser fracasos para que el primer éxito ocurra en el ensayo k . Además, como los eventos son independientes y en cada ensayo la probabilidad de éxito es p y la de fracaso es $1 - p$, resulta que

$$(2.3) \quad P(Y = k) = P(A_1^c)P(A_2^c) \cdots P(A_{k-1}^c)P(A_k) = (1 - p)^{k-1}p, \quad k = 1, 2, \dots$$

Se dice que la variable aleatoria Y tiene distribución geométrica de parámetro p y se denota

$$Y \sim \mathcal{G}(p).$$

Así, el número de tiradas de una moneda hasta obtener cara y el número de tiradas de un dado hasta obtener un 6 tienen distribución geométrica de parámetros $\frac{1}{2}$ y $\frac{1}{6}$, respectivamente.

Muchos autores suelen llamar distribución geométrica a la correspondiente a la variable X que cuenta el número de *fracasos* hasta que se produce el primer éxito, en vez del número total de ensayos. En esencia, estamos hablando de la misma variable salvo por un corrimiento en una constante. Esto es, $X = Y - 1$ y

$$P(X = k) = P(Y = k + 1) = (1 - p)^k p$$

para $k = 0, 1, 2, \dots$. Esta es la definición que utiliza R en sus funciones `dgeom` y `pgeom` para el cálculo de las funciones de probabilidad y de distribución, respectivamente. A los fines de los resultados que se quieren ilustrar en este artículo, adoptaremos la definición correspondiente a (2.3).

Una variable aleatoria con distribución geométrica se caracteriza por cumplir la siguiente propiedad:

Propiedad de falta de memoria de la distribución geométrica: Sea $Y \sim \mathcal{G}(p)$, entonces

$$P(Y > n + m \mid Y > n) = P(Y > m) \quad \text{para todo } n, m \in \mathbb{N}_0.$$

Esto significa que si hasta el ensayo n no ocurrió un éxito (esto es, $Y > n$), la probabilidad de que no haya un éxito en los siguientes m ensayos (hasta el momento

$n + m$) es la misma que si se iniciara nuevamente la secuencia de ensayos y no ocurriera un éxito en los primeros m ensayos. Esto se ve claro cuando arrojamus una moneda hasta obtener una cara o de un dado hasta obtener el número 6. Las nuevas tiradas no “recuerdan” los resultados en una tanda de tiradas anterior. Por eso decimos que la distribución geométrica “no tiene memoria”.

Demostración: Notemos primero que para todo $n \in \mathbb{N}_0$ se cumple $P(Y > n) = (1 - p)^n$ ya que, para $n = 0$, $P(Y > 0) = 1 = (1 - p)^0$ y, si $n > 0$,

$$\begin{aligned} P(Y > n) &= 1 - P(Y \leq n) = 1 - \sum_{y=1}^n (1-p)^{y-1} p = 1 - p \sum_{i=0}^{n-1} (1-p)^i \\ &= 1 - p \frac{(1-p)^n - 1}{-p} = (1-p)^n. \end{aligned}$$

Luego,

$$\begin{aligned} P(Y > n + m \mid Y > n) &= \frac{P(\{Y > n + m\} \cap \{Y > n\})}{P(Y > n)} \\ (2.4) \qquad &= \frac{P(Y > n + m)}{P(Y > n)} \\ &= \frac{(1-p)^{n+m}}{(1-p)^n} = (1-p)^m = P(Y > m), \end{aligned}$$

como queríamos demostrar. \square

La propiedad de falta de memoria también puede escribirse como

$$P(Y > n + m) = P(Y > n)P(Y > m)$$

para todo $n, m \in \mathbb{N}_0$, lo cual puede verse de (2.4), despejando $P(Y > n + m)$. Un resultado notable es que la distribución geométrica es la única distribución discreta, con valores posibles el conjunto de los números naturales, que cumple con la propiedad de falta de memoria. Esto es consecuencia del siguiente resultado:

Proposición (Caracterización de la distribución geométrica): Sea Y una variable aleatoria discreta con valores posibles $1, 2, \dots$ tal que

$$P(Y > n + m) = P(Y > n)P(Y > m), \quad \text{para todo } n, m \in \mathbb{N}_0.$$

Entonces, Y tiene distribución geométrica.

Demostración: La demostración de este resultado se encuentra en el Apéndice. \square

Otro ejemplo de esta situación ocurre cuando una pieza de un equipo es observada durante períodos fijos de tiempo (como horas o días) para ver si ha fallado, registrándose si falla en la primera hora (o día), en la segunda, etc. (Hoel y cols., 1971). Si la pieza no se deteriora ni mejora con el transcurso del tiempo (por ejemplo, un motor que recibe mantenimiento periódico), y puede fallar debido a causas

esporádicas que ocurren homogéneamente en el tiempo, entonces es razonable pensar que el momento en que la pieza falla cumple con la propiedad de falta de memoria y, por lo tanto, tiene distribución geométrica.

2.6. La distribución binomial negativa. Para finalizar nuestro recorrido por las distribuciones de variables aleatorias discretas, estudiaremos la distribución binomial negativa. Surge en el mismo contexto de la distribución geométrica, pero ahora nuestra variable Y cuenta el número total de ensayos que ocurren hasta que se obtienen r éxitos. Así, la distribución geométrica resulta un caso particular de la distribución binomial negativa, para $r = 1$. Notemos que ahora los valores posibles de Y son los naturales mayores o iguales a r , ya que al menos se necesitan r ensayos para alcanzar los r éxitos (que sería el caso en que los primeros r ensayos son todos exitosos), pero el número total de ensayos podría crecer indefinidamente. ¿Cómo calculamos entonces $P(Y = k)$, para $k = r, r + 1, \dots$? Primero notemos que un punto cualquiera en el evento $\{Y = k\}$ contiene r éxitos y $k - r$ fracasos y, por lo tanto, como los ensayos son independientes, tiene probabilidad $p^r(1 - p)^{k-r}$. Pero, ¿cuántos puntos muestrales hay en este evento? Como necesariamente el último de los k ensayos es un éxito (porque al lograr el éxito r se finaliza el experimento), hay tantos puntos como formas de ubicar los restantes $r - 1$ éxitos en los primeros $k - 1$ lugares, esto es, el número combinatorio $\binom{k-1}{r-1}$. También podemos pensar en ubicar los $k - r$ fracasos en los primeros $k - 1$ lugares, lo cual corresponde al número combinatorio $\binom{k-1}{k-r}$, que, por definición, coincide con $\binom{k-1}{r-1}$. Así,

$$P(Y = k) = \binom{k-1}{r-1} p^r (1 - p)^{k-r}, \quad \text{para } k = r, r + 1, \dots$$

Se dice que Y tiene distribución binomial negativa de parámetros r y p y se denota

$$Y \sim BN(r, p).$$

Ejemplo: Se sabe por estudios previos que el 30 % de una población está de acuerdo con una nueva medida del gobierno. Si se realiza una encuesta telefónica preguntando al encuestado si acuerda o no con la medida y la selección de los encuestados se realiza al azar, ¿cuál es la probabilidad de que el encuestado en la séptima llamada sea el tercero en estar de acuerdo? ¿Y de que se tengan que realizar a lo sumo 7 llamadas para encontrar a 3 personas que estén de acuerdo?

Solución: Si la población es suficientemente grande, con argumentos similares a los ya utilizados para la distribución binomial, podemos asumir que cada llamada puede resultar en éxito (está de acuerdo) o fracaso (no está de acuerdo) con probabilidad 0,3 o 0,7, respectivamente, independientemente de las demás llamadas. Así, la variable Y que cuenta el número de llamadas que se deben realizar hasta encontrar a 3 personas de acuerdo con la nueva medida tiene distribución binomial negativa con parámetros $r = 3$ y $p = 0,3$. Luego, la probabilidad de que el

encuestado en la séptima llamada sea el tercero en estar de acuerdo corresponde a

$$P(Y = 7) = \binom{6}{2} 0,3^3 \times 0,7^4 = 0,0972.$$

Y la probabilidad de que se tengan que realizar a lo sumo 7 llamadas para encontrar a 3 personas que estén de acuerdo es

$$P(Y \leq 7) = \sum_{k=3}^7 \binom{k-1}{2} 0,3^3 \times 0,7^{k-3} = 0,3529.$$

Como ya se comentó para la distribución geométrica, muchos autores consideran que la distribución binomial negativa corresponde a la variable X que cuenta el número de *fracasos* hasta que ocurren los r éxitos, en lugar del número *total* de ensayos. Ambas variables se relacionan mediante $X = Y - r$, de manera que X toma valores en \mathbb{N}_0 . Así, para $k = 0, 1, 2, \dots$,

$$P(X = k) = P(Y = k + r) = \binom{k+r-1}{r-1} p^r (1-p)^k.$$

Esta es la definición adoptada por el software R. Así, en el ejemplo de recién, $P(Y = 7) = P(X = 4)$ y puede obtenerse con el comando `dnbinom(4, 3, 0.3)`, mientras que $P(Y \leq 7) = P(X \leq 4)$ y se obtiene mediante `pnbinom(4, 3, 0.3)`.

§3. Variables aleatorias continuas

Dijimos que una variable continua es aquella que puede tomar cualquier valor en un intervalo. Formalmente, una variable aleatoria Y es continua si satisface

$$P(Y = y) = 0$$

para todo y en \mathbb{R} . Esto trae como consecuencia que Y es continua si y sólo si su función de distribución F es continua.

Para la mayoría de las variables aleatorias continuas, F puede escribirse en términos de lo que se conoce como función de densidad f . Esto es,

$$F(y) = \int_{-\infty}^y f(t) dt,$$

donde f satisface $f(y) \geq 0$ para todo y en \mathbb{R} y $\int_{-\infty}^{\infty} f(y) dy = 1$. A diferencia del caso discreto, la función de densidad no corresponde a la probabilidad de que la variable aleatoria tome un cierto valor, ya que esa probabilidad siempre es cero, sino que nos permite calcular, por ejemplo, la probabilidad de caer en un intervalo. Así, la probabilidad de que Y tome valores en un intervalo (a, b) puede calcularse como

$$P(a < Y < b) = \int_a^b f(y) dy,$$

que corresponde al área bajo la curva entre los valores a y b . Como la variable toma valores en \mathbb{R} , la probabilidad de caer en $(-\infty, \infty)$ tiene necesariamente probabilidad 1 y es por eso que el área total bajo la curva es siempre 1 ($\int_{-\infty}^{\infty} f(y)dy = 1$).

Si conocemos la función de distribución F y esta es tal que F' existe y es continua, excepto quizás en un número finito de puntos, podemos obtener una función de densidad f como

$$(3.1) \quad f(y) = \begin{cases} F'(y), & \text{en los puntos donde } F' \text{ existe y es continua,} \\ 0, & \text{caso contrario.} \end{cases}$$

En realidad no debemos preocuparnos por la definición de f en un número finito de puntos, ya que el cambio en el valor de la densidad en esos puntos no modifica la función de distribución de la variable aleatoria.

El valor esperado de una variable aleatoria continua Y con función de densidad f se define como

$$E(Y) = \int_{-\infty}^{\infty} yf(y)dy$$

si $\int_{-\infty}^{\infty} |y|f(y)dy < \infty$. Al igual que para las variables aleatorias discretas, la varianza de Y se define según (2.1). Para el cálculo, nuevamente es útil la expresión equivalente $Var(Y) = E(Y^2) - E(Y)^2$, donde

$$E(Y^2) = \int_{-\infty}^{\infty} y^2 f(y)dy.$$

3.1. La distribución uniforme. Para iniciar nuestro recorrido por las distribuciones de variables aleatorias continuas, estudiaremos una de las más simples: la distribución uniforme. Surge, por ejemplo, cuando elegimos un punto al azar en el intervalo (a, b) . No le damos mayor peso a un resultado sobre otro, por lo cual es igualmente probable estar cerca de cualquier punto en (a, b) (Ross, 2014). Así, la densidad es constante en el intervalo (a, b) y vale 0 fuera de ese intervalo, es decir,

$$f(y) = \begin{cases} k, & a < y < b, \\ 0, & \text{caso contrario.} \end{cases}$$

¿Cómo encontramos el valor de k ? Sabemos que el área total bajo la curva debe ser 1 y, como $f(y) = 0$ fuera del intervalo (a, b) , resulta $1 = \int_a^b k dy = k(b - a)$. Luego, $k = \frac{1}{b-a}$. Decimos entonces que una variable Y con función de densidad f dada por

$$f(y) = \begin{cases} \frac{1}{b-a}, & a < y < b, \\ 0, & \text{caso contrario.} \end{cases}$$

tiene distribución uniforme en (a, b) y lo denotamos

$$Y \sim \mathcal{U}(a, b).$$

El intervalo también puede considerarse cerrado en los extremos.

En la distribución uniforme, la probabilidad de caer en un intervalo totalmente contenido en (a, b) no depende de la ubicación de dicho intervalo sino de su *longitud*. Es decir, si (c, d) y (e, f) son dos intervalos contenidos en (a, b) de igual longitud, digamos L (esto es $d - c = L = f - e$), entonces la probabilidad de caer en cualquiera de estos intervalos es la misma y vale

$$\int_c^d \frac{1}{b-a} dy = \int_e^f \frac{1}{b-a} dy = \frac{L}{b-a}.$$

Ejemplo (Canavos, 1995): Si el peso de un individuo se redondea al kilogramo más cercano, entonces la diferencia entre este y el peso verdadero será un valor entre $-0,5$ y $0,5$ kg y podemos pensar que el error de redondeo se distribuye uniformemente en el intervalo $(-0,5; 0,5)$.

Ilustraremos para el caso de una variable uniforme el cálculo de esperanza y varianza de una variable aleatoria continua.

Proposición: Sea $Y \sim \mathcal{U}(a, b)$. Entonces, $E(Y) = \frac{a+b}{2}$ y $Var(Y) = \frac{(b-a)^2}{12}$.

Demostración: Si $Y \sim \mathcal{U}(a, b)$, entonces el valor esperado de Y se obtiene mediante

$$E(Y) = \int_a^b \frac{y}{b-a} dy = \frac{1}{b-a} \frac{y^2}{2} \Big|_a^b = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{1}{b-a} \frac{(b-a)(b+a)}{2} = \frac{a+b}{2},$$

es decir, es el punto medio del intervalo (a, b) .

Para calcular la varianza de Y , notemos primero que

$$E(Y^2) = \int_a^b y^2 \frac{1}{b-a} dy = \frac{1}{b-a} \frac{y^3}{3} \Big|_a^b = \frac{b^3 - a^3}{3(b-a)}.$$

Así,

$$Var(Y) = \frac{b^3 - a^3}{3(b-a)} - \frac{(a+b)^2}{4} = \frac{4b^3 - 4a^3 - 3(a+b)(b^2 - a^2)}{12(b-a)} = \frac{b^3 - a^3 - 3ab^2 + 3ba^2}{12(b-a)} = \frac{(b-a)^3}{12(b-a)} = \frac{(b-a)^2}{12},$$

como queríamos demostrar. \square

La distribución uniforme es especialmente útil para la simulación de valores de una variable aleatoria con una distribución específica (Hoel y cols., 1971; Ross, 1999; Canavos, 1995). Una forma de hacerlo se basa en el siguiente resultado:

Proposición: Sea $U \sim \mathcal{U}(0, 1)$ y sea F una función de distribución continua con inversa F^{-1} . Entonces, la variable aleatoria Y definida como $Y = F^{-1}(U)$ tiene distribución F .

Demostración: Notemos primero que la variable U toma valores en $(0, 1)$ con probabilidad 1. Además, $F^{-1}(U)$ está bien definida pues $F^{-1}(y)$ tiene que estar definida para todo y en $(0, 1)$ por las propiedades que cumple en general una función de distribución ($\lim_{y \rightarrow -\infty} F(y) = 0$ y $\lim_{y \rightarrow \infty} F(y) = 1$) y por ser en este caso F continua

(teorema del valor intermedio). Además, F es estrictamente creciente, por ser en general no decreciente y por existir F^{-1} (F es inyectiva).

Veamos cuál es la función de distribución de Y . Sea y en \mathbb{R} , entonces

$$F_Y(y) = P(Y \leq y) = P(F^{-1}(U) \leq y) = P(U \leq F(y))$$

por ser F creciente (y por lo tanto $a \leq b$ si y sólo si $F(a) \leq F(b)$) y $F(F^{-1}(U)) = U$. Así, como $F(y)$ está en $(0, 1)$, resulta

$$P(U \leq F(y)) = \int_{-\infty}^{F(y)} f_U(u) du = \int_0^{F(y)} 1 du = F(y).$$

Es decir, Y tiene distribución F , como queríamos demostrar. \square

De esta manera, si podemos generar un número aleatorio entre 0 y 1, podemos generar un valor de una variable aleatoria con distribución F aplicando la inversa de la función al número generado.

3.2. La distribución exponencial. La distribución exponencial se utiliza frecuentemente para modelar tiempos de espera hasta que ocurre un determinado suceso, tiempos de sobrevivencia, duración de artículos. ¿Por qué es razonable esto? Pensemos a modo de ejemplo en la variable aleatoria T que mide el tiempo que transcurre hasta que ocurre la primera llamada en una central telefónica. Si λ es el número medio de llamadas por unidad de tiempo, es razonable asumir que el número de llamadas que ocurre en un intervalo de tiempo de longitud t tiene distribución $\mathcal{P}(\lambda t)$. Busquemos ahora la función de distribución de T . Como T sólo toma valores positivos, es claro que $P(T \leq t) = 0$ si $t \leq 0$. ¿Pero qué ocurre para un tiempo t positivo? Notemos que el tiempo que transcurre hasta que ocurre la primera llamada es mayor a t si y sólo si hasta el tiempo t no se recibió ninguna llamada, es decir, el número de llamadas N_t recibidas en el intervalo $[0, t]$ es 0. En fórmulas,

$$P(T \leq t) = 1 - P(T > t) = 1 - P(N_t = 0) = 1 - e^{-\lambda t},$$

pues $N_t \sim \mathcal{P}(\lambda t)$. Así, la función de distribución acumulada F de T es

$$F(t) = \begin{cases} 1 - e^{-\lambda t}, & t > 0, \\ 0, & t \leq 0. \end{cases}$$

Si ahora derivamos F para encontrar la densidad f de T , como en (3.1), llegamos a la siguiente expresión:

$$(3.2) \quad f(t) = \begin{cases} \lambda e^{-\lambda t}, & t > 0, \\ 0, & t \leq 0. \end{cases}$$

Decimos que una variable aleatoria Y tiene distribución exponencial de parámetro λ , con $\lambda > 0$, y se denota

$$Y \sim \text{Exp}(\lambda),$$

si su función de densidad está dada por (3.2). Así, la variable T que mide el tiempo hasta que ocurre la primera llamada en una central telefónica tiene distribución exponencial.

Ejemplo: ¿Cuál es la probabilidad de que en una central telefónica se reciba la primera llamada después de media hora, si se reciben en promedio 6 llamadas por hora?

Solución: Si consideramos la variable T que mide el tiempo (en horas) hasta que ocurre la primera llamada, entonces $T \sim \text{Exp}(6)$. Notemos que si $Y \sim \text{Exp}(\lambda)$, podemos deducir que $P(Y > y) = 1 - F(y) = e^{-\lambda y}$, si $y \geq 0$. Así, la probabilidad buscada es $P(T > \frac{1}{2}) = e^{-6 \cdot \frac{1}{2}} = 0,0498$.

La distribución exponencial es la análoga continua a la distribución geométrica, ya que posee también la propiedad de falta de memoria (Wackerly y cols., 2010; Bertsekas y Tsitsiklis, 2008; Hoel y cols., 1971).

Propiedad de falta de memoria de la distribución exponencial: Sea $Y \sim \text{Exp}(\lambda)$, entonces

$$(3.3) \quad P(Y > a + b \mid Y > a) = P(Y > b) \quad \text{para todo } a, b \geq 0.$$

Demostración: Al ser $a, b \geq 0$, resulta $\{Y > a + b\} \cap \{Y > a\} = \{Y > a + b\}$ y

$$\begin{aligned} P(Y > a + b \mid Y > a) &= \frac{P(\{Y > a + b\} \cap \{Y > a\})}{P(Y > a)} = \frac{P(Y > a + b)}{P(Y > a)} \\ &= \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}} = e^{-\lambda b} = P(Y > b) \end{aligned}$$

como se quería ver. □

Esta propiedad también puede escribirse como

$$P(Y > a + b) = P(Y > a)P(Y > b) \quad \text{para todo } a, b \geq 0,$$

lo cual se deduce de los pasos de la demostración.

La interpretación de este resultado es similar a la que dimos para la distribución geométrica: si pensamos por ejemplo en la vida útil de una pieza (como un componente eléctrico), (3.3) nos dice que la probabilidad de una vida útil *adicional* en más de b unidades es la misma que la probabilidad *original* de vida útil en más de b unidades. En este caso, la antigüedad de la pieza no aumenta ni disminuye la probabilidad de falla en un período de tiempo determinado (Hoel y cols., 1971).

La propiedad de falta de memoria es muy importante ya que caracteriza a la distribución exponencial, en el sentido de que una variable continua y positiva que cumple con esa propiedad, necesariamente tiene que tener distribución exponencial. Esto nos lo dice el siguiente resultado:

Proposición (Caracterización de la distribución exponencial): Sea Y una variable aleatoria tal que

$$P(Y > a + b) = P(Y > a)P(Y > b) \quad \text{para todo } a, b \geq 0.$$

Entonces, $P(Y > 0) = 0$ o Y tiene distribución exponencial.

Demostración: La demostración de este resultado se encuentra en el Apéndice. \square

Otra relación interesante entre la distribución exponencial y la geométrica es la siguiente. Consideremos una variable aleatoria T con distribución exponencial de parámetro λ . Como ya vimos, T puede indicar el tiempo que transcurre hasta que falla cierto componente. Pensemos ahora en la “versión discreta” Y de T , subdividiendo el tiempo en intervalos de longitud uno y considerando en cuál de estos intervalos ocurrió el evento. Por ejemplo, si el tiempo se mide en horas, $Y = 1$ significa que la falla ocurrió durante la primera hora y, en general, $Y = k$ significa que la falla ocurrió en la hora k -ésima. La variable aleatoria Y claramente toma solamente valores naturales por lo cual es una variable aleatoria discreta. Veamos que Y tiene distribución geométrica. Para $k = 1, 2, \dots$, la falla ocurre durante la k -ésima hora si el tiempo hasta que ocurre la falla, medido en escala continua, está entre $k - 1$ y k . Así,

$$\begin{aligned} P(Y = k) &= P(k - 1 < T \leq k) = F_T(k) - F_T(k - 1) \\ &= 1 - e^{-\lambda k} - (1 - e^{-\lambda(k-1)}) = e^{-\lambda(k-1)}(1 - e^{-\lambda}) = (1 - p)^{k-1}p, \end{aligned}$$

para $p = 1 - e^{-\lambda} = P(T \leq 1)$ (que cumple $0 < p < 1$). Por lo tanto, Y tiene distribución geométrica de parámetro $p = 1 - e^{-\lambda}$.

Formalmente, la variable Y es el “techo” de T , y se denota $\lceil T \rceil$, donde el techo de un número x corresponde al menor entero mayor o igual a x . Luego, acabamos de ver la demostración del siguiente resultado:

Proposición: Si T es una variable aleatoria tal que $T \sim \text{Exp}(\lambda)$, entonces se tiene $\lceil T \rceil \sim \mathcal{G}(1 - e^{-\lambda})$.

3.3. La distribución Gamma. Esta es una distribución que sirve para modelar variables aleatorias continuas que asumen valores positivos, generalmente asimétricas. Decimos que una variable aleatoria tiene distribución Gamma con parámetros α y λ , y lo denotamos como

$$Y \sim \Gamma(\alpha, \lambda)$$

si la función de densidad está dada por

$$(3.4) \quad f(y) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y}, & y > 0, \\ 0, & y \leq 0, \end{cases}$$

con $\alpha, \lambda > 0$ y $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$. A α se lo conoce como *parámetro de forma*, ya que justamente tiene que ver con la forma de la gráfica de la función de densidad:

cuando y tiende a 0 por derecha, $f(y)$ tiende a λ si $\alpha = 1$, a 0 si $\alpha > 1$ y a ∞ si $\alpha < 1$. Además, la función es decreciente en $(0, \infty)$ para $\alpha \leq 1$, y para $\alpha > 1$ crece desde 0 hasta alcanzar un máximo y luego decrece. El parámetro λ tiene que ver con la escala (siendo $\beta = \frac{1}{\lambda}$ conocido como *parámetro de escala*), ya que la gráfica se comprime o se estira dependiendo de su valor. La Figura 2 presenta distintas gráficas de la densidad Gamma, como así también de la correspondiente función de distribución, variando los valores de α y λ .

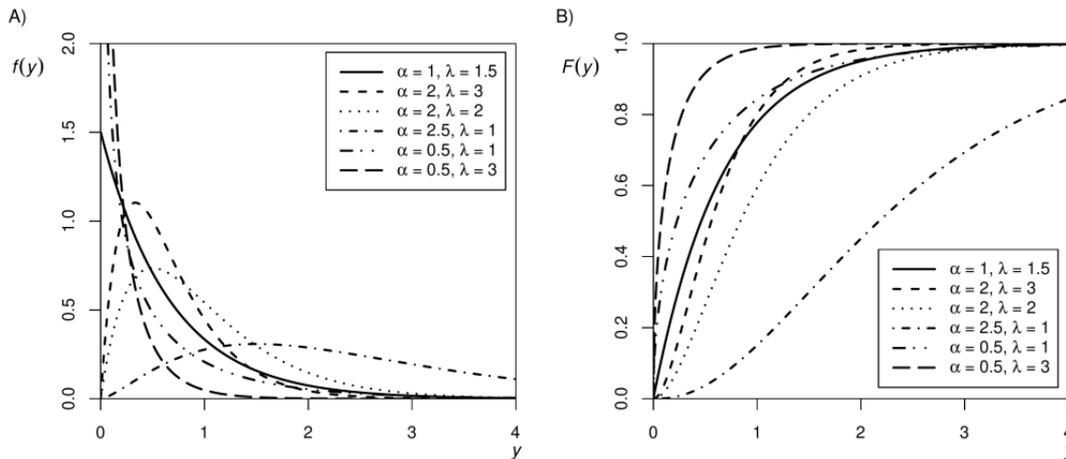


FIGURA 2. Gráficas de la función de densidad (A) y de la función de distribución (B) de una variable aleatoria con distribución Gamma, para distintos valores de los parámetros α y λ . En todos los casos $f(y) = 0 = F(y)$ para $y < 0$.

La distribución exponencial de parámetro λ es un caso particular de la distribución Gamma para $\alpha = 1$. También se asocia la distribución Gamma a tiempos de sobrevivida, tiempos de espera. Recordando el ejemplo que vimos para introducir a la distribución exponencial, podemos llegar a la expresión (3.4) para $\alpha \in \mathbb{N}$, pensando en el tiempo hasta que ocurren n llamadas (o en general, sucesos de algún tipo, accidentes, etc.). En este caso particular, la distribución Gamma también se conoce como distribución de Erlang. Para ver esto, consideremos las siguientes variables aleatorias:

- N_t = número de sucesos que ocurren en el intervalo $[0, t]$,
- T_n = tiempo de espera hasta que ocurren n sucesos.

Luego, como antes, $N_t \sim \mathcal{P}(\lambda t)$ y $P(T_n \leq t) = 0$ si $t \leq 0$. Para $t > 0$, la clave está en notar lo siguiente: que el tiempo de espera hasta que ocurran n sucesos sea a lo sumo t se corresponde directamente con el hecho de que ocurran por lo menos n sucesos (n o más) en el intervalo $[0, t]$. Menos de n no podrían suceder pues, en ese caso, habría que esperar más tiempo. Pero podrían ocurrir más de n , ya que si ocurrieron n en un tiempo menor estricto a t , es posible que haya ocurrido algún

suceso más hasta el tiempo t . En términos probabilísticos,

$$F_{T_n}(t) = P(T_n \leq t) = P(N_t \geq n) = \sum_{j=n}^{\infty} P(N_t = j) = \sum_{j=n}^{\infty} \frac{e^{-\lambda t} (\lambda t)^j}{j!}.$$

Ahora debemos derivar $F_{T_n}(t)$ respecto de t para obtener la densidad $f_{T_n}(t)$, que vale 0 para $t < 0$ y, para $t > 0$,

$$\begin{aligned} f_{T_n}(t) &= -\lambda e^{-\lambda t} \sum_{j=n}^{\infty} \frac{(\lambda t)^j}{j!} + e^{-\lambda t} \sum_{j=n}^{\infty} \frac{j(\lambda t)^{j-1} \lambda}{j!} \\ &= -\lambda e^{-\lambda t} \sum_{j=n}^{\infty} \frac{(\lambda t)^j}{j!} + \lambda e^{-\lambda t} \sum_{j=n}^{\infty} \frac{(\lambda t)^{j-1}}{(j-1)!}. \end{aligned}$$

Separando en la segunda sumatoria el término correspondiente a $j = n$ y tomando $i = j - 1$, resulta

$$\begin{aligned} f_{T_n}(t) &= -\lambda e^{-\lambda t} \sum_{j=n}^{\infty} \frac{(\lambda t)^j}{j!} + \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} + \lambda e^{-\lambda t} \sum_{i=n}^{\infty} \frac{(\lambda t)^i}{i!} \\ &= \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} = \frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t}, \end{aligned}$$

usando el resultado de que $\Gamma(n) = (n-1)!$ para $n \in \mathbb{N}$. Esto es, $T_n \sim \Gamma(n, \lambda)$. Así, hemos probado la proposición que se enuncia a continuación:

Proposición: Si el número de sucesos que ocurren en el intervalo $[0, t]$ tiene distribución $\mathcal{P}(\lambda t)$ para todo $t > 0$, entonces el tiempo de espera hasta que ocurren n sucesos tiene distribución Gamma con parámetros n y λ .

Para ilustrar este resultado veamos el siguiente ejemplo:

Ejemplo: Supongamos que, en promedio, el número de personas que entran a un supermercado por minuto es de 5, siguiendo una distribución de Poisson. ¿Cuál es la probabilidad de que en menos de 20 segundos entren dos personas?

Solución: Podemos considerar que el número de personas que entran al supermercado en un intervalo de tiempo (medido en minutos) de longitud t tiene distribución $\mathcal{P}(5t)$. Luego, por el resultado anterior, el tiempo T que hay que esperar hasta que entren dos personas tiene distribución $\Gamma(2, 5)$. Estamos interesados en calcular la probabilidad de que T sea menor a 20 segundos. Como 20 segundos corresponden a $1/3$ de minuto, la probabilidad buscada es

$$P(T < \frac{1}{3}) = \int_0^{\frac{1}{3}} \frac{5^2}{\Gamma(2)} t e^{-5t} dt = 5(-e^{-5t}(t + \frac{1}{5})) \Big|_0^{\frac{1}{3}} = 0,4963.$$

3.4. La distribución normal. Para finalizar nuestro recorrido por las distintas distribuciones de probabilidad, no podemos dejar de mencionar a la distribución

normal, que es la más importante, y es ampliamente utilizada para describir distintos fenómenos que ocurren en la práctica. Conozcamos un poco de su historia (Devore, 2005; Canavos, 1995; Ross, 2014): Esta distribución fue descubierta por el matemático francés Abraham DeMoivre en 1733 como límite de la distribución binomial con $p = \frac{1}{2}$, resultado que fue extendido por Laplace para un p general en 1812. Este es un caso particular de un resultado clave para la inferencia estadística, el *Teorema Central del Límite*, que plantea que, bajo ciertas condiciones, la suma (y el promedio) de un número grande de variables aleatorias tiene una distribución aproximadamente normal. También se la llama distribución gaussiana, o campana de Gauss a la densidad (por su forma acampanada), a partir de que Gauss la mencionó en un artículo en 1809, utilizándola para predecir la ubicación de objetos astronómicos. Entre mediados y fines del siglo XIX, gran parte de los estadísticos comenzaron a creer que la mayoría de los conjuntos de datos tendrían histogramas siguiendo esta forma acampanada y que era “normal” que datos que se comportaran razonablemente siguieran esta curva; de allí se la empezó a llamar “curva normal”.

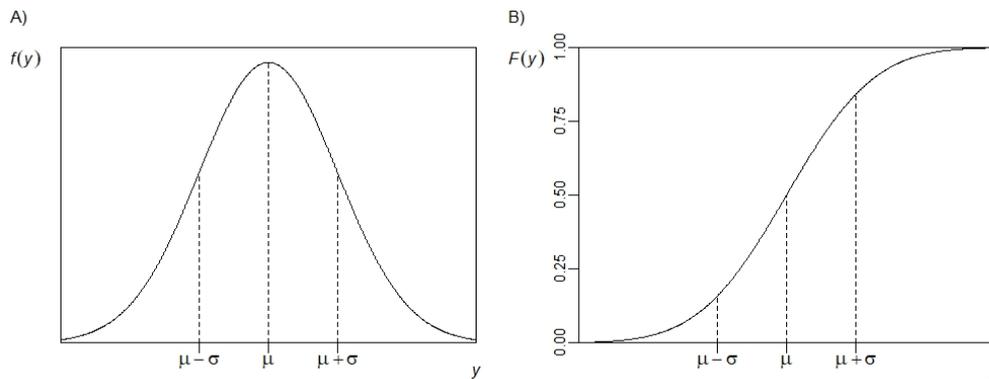


FIGURA 3. Gráficas de la función de densidad (A) y la función de distribución (B) de una variable aleatoria con distribución $N(\mu, \sigma^2)$.

La función de densidad de una variable Y con distribución normal de parámetros μ y σ^2 , con $\mu \in \mathbb{R}$ y $\sigma > 0$, está dada por

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \quad y \in \mathbb{R}.$$

Se denota

$$Y \sim N(\mu, \sigma^2).$$

Si bien una deducción de esta expresión está fuera del alcance de este artículo, mencionaremos algunos aspectos importantes de esta distribución.

La densidad normal tiene una forma acampanada como se muestra en la Figura 3 A), tiende a 0 tanto en la dirección positiva como negativa sin alcanzar nunca ese valor, está centrada en μ (es simétrica respecto a μ), siendo mayor la

probabilidad de obtener valores en zonas alejadas de μ a medida que σ aumenta. Puede mostrarse que μ y σ^2 corresponden a la media y la varianza de la distribución, respectivamente (ver, por ejemplo, Hoel y cols. (1971)). En la Figura 3 B) se observa la función de distribución acumulada, que es continua en todo punto por ser Y una variable aleatoria continua.

En la práctica, es razonable intuir que un cierto conjunto de datos podría ser modelado por una distribución normal si al graficar un histograma de frecuencias observamos simetría y podemos trazar una curva suave con la forma acampanada de la densidad normal. Ejemplos de variables que suelen tener distribución aproximadamente normal son el peso, la altura, la temperatura, errores en mediciones experimentales, calificaciones en una prueba, etc.

§4. Comentarios finales

En este artículo hemos emprendido un recorrido por las distribuciones de probabilidad más conocidas, que puede ser de utilidad para quien esté transitando un curso de Probabilidad o para aportar elementos a quien tenga estudios previos, brindando quizás algunas ideas nuevas y contribuyendo a cerrar conceptos. Las Tablas 3 y 4 presentan una síntesis de lo estudiado, incluyendo la expresión de la media y la varianza para cada una de las distribuciones.

El estudio de la Probabilidad resulta de suma importancia ya que los conceptos de probabilidad constituyen la base para poder realizar Inferencia Estadística, es decir, obtener conclusiones sobre características de una población a partir de una muestra extraída de esa población.

En la práctica, contamos con datos, que corresponden al resultado de observaciones efectuadas de una variable Y (característica de interés) en los individuos que conforman la muestra. También podríamos querer explicar una variable de respuesta Y en términos de otras variables X_1, \dots, X_p medidas en la muestra, lo cual es el objetivo de los modelos de regresión. Para el análisis, en general se requiere asumir una distribución para la variable aleatoria Y . En función de la naturaleza de la variable (continua, discreta) y de las condiciones experimentales y características de la medición (si sólo toma valores positivos, si corresponde a tiempos de espera, si es un experimento binomial, etc.) podemos proponer una distribución que resulte razonable para Y . Los parámetros de dicha distribución podrán ser estimados a partir de los valores observados de la variable en la muestra, para lo cual existen métodos de estimación. Por ejemplo, la probabilidad de éxito p de una binomial podrá estimarse a partir de la proporción de individuos en la muestra que posean la característica considerada como éxito.

Este trabajo intenta mejorar la comprensión de las distribuciones de probabilidad más utilizadas. Conocer en mayor profundidad bajo qué condiciones surge cada

una de ellas nos ayudará a tomar mejores decisiones al momento de atribuir una distribución posible a una variable de interés.

Tabla 3: Resumen de las distribuciones de probabilidad estudiadas para variables aleatorias discretas.

Distribución	$f(y)$	Usos y características	Media	Varianza
Binomial $B(n, p)$	$\binom{n}{y} p^y (1-p)^{n-y}$, $y = 0, 1, \dots, n$	Número de éxitos entre n ensayos independientes con la misma probabilidad p de éxito. Aproxima a la distribución hipergeométrica $\mathcal{H}(n, M, N)$ cuando n es pequeño en relación a N .	np	$np(1-p)$
Hipergeométrica $\mathcal{H}(n, M, N)$	$\frac{\binom{M}{y} \binom{N-M}{n-y}}{\binom{N}{n}}$, máx $\{0, n-N+M\}$ $\leq y \leq \min \{n, M\}$, $y \in \mathbb{Z}$	Número de objetos de tipo I en una muestra de tamaño n extraída sin reposición de una población con N objetos, de los cuales M son de tipo I y el resto de tipo II.	$n \frac{M}{N}$	$n \frac{M}{N} \frac{N-M}{N} \frac{N-n}{N-1}$
Poisson $\mathcal{P}(\lambda)$	$\frac{e^{-\lambda} \lambda^y}{y!}$, $y = 0, 1, \dots$	Número de sucesos que ocurren en un determinado tiempo o espacio. Aproxima a la distribución binomial $B(n, p)$ para n grande, p pequeña y $\lambda = np$ moderado.	λ	λ
Geométrica $\mathcal{G}(p)$	$(1-p)^{y-1} p$, $y = 1, 2, \dots$	Número de ensayos independientes hasta que ocurre el primer éxito, con probabilidad p de éxito en cada ensayo. Cumple con la propiedad de falta de memoria.	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Binomial negativa $BN(r, p)$	$\binom{y-1}{r-1} p^r (1-p)^{y-r}$, $y = r, r+1, \dots$	Número de ensayos independientes hasta que ocurren r éxitos, con probabilidad p de éxito en cada ensayo.	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$

Nota: En todos los casos se indican los valores de y para los cuales $f(y) \neq 0$.

Tabla 4: Resumen de las distribuciones de probabilidad estudiadas para variables aleatorias continuas.

Distribución	$f(y)$	Usos y características	Media	Varianza
Uniforme $\mathcal{U}(a, b)$	$\frac{1}{b-a},$ $a < y < b$	Punto elegido al azar en el intervalo (a, b) . La probabilidad de caer en un intervalo contenido en (a, b) depende sólo de su longitud y no de su ubicación.	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponencial $Exp(\lambda)$	$\lambda e^{-\lambda y},$ $y > 0$	Tiempo de espera hasta que ocurre un suceso, tiempo de sobrevida, duración de artículos. Cumple con la propiedad de falta de memoria.	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma $\Gamma(\alpha, \lambda)$	$\frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y},$ $y > 0$	Variables positivas, con distribución generalmente asimétrica (tiempo de sobrevida, tiempo de espera).	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
Normal $N(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}},$ $y \in \mathbb{R}$	Variables con densidad simétrica, de forma a-campañada (peso, altura, temperatura, errores de medición, calificaciones, etc.).	μ	σ^2

Nota: En todos los casos se indican los valores de y para los cuales $f(y) \neq 0$.

§5. Apéndice

Otra manera de deducir la distribución de Poisson: La distribución de Poisson se utiliza en situaciones donde se quiere estudiar la ocurrencia esporádica de ciertos fenómenos o sucesos a lo largo del tiempo, como por ejemplo, la aparición de terremotos, el ingreso de clientes en un banco, la ocurrencia de accidentes. Bajo ciertas suposiciones, veremos que esta distribución modeliza el número de veces que ocurre el fenómeno bajo estudio en el período de tiempo considerado. Supongamos que se cumplen las siguientes condiciones:

- (1) La probabilidad de que exactamente un suceso ocurra en un intervalo de tiempo de longitud h es la misma para cualquier intervalo de longitud h y es igual a $\lambda h + o(h)$, donde $\lambda > 0$ y $o(h)$ representa cualquier función $f(h)$ tal que $\lim_{h \rightarrow 0} \frac{f(h)}{h} = 0$.
- (2) La probabilidad de que ocurran dos o más sucesos en un intervalo de longitud h es la misma para cualquier intervalo de longitud h y es igual a $o(h)$.

- (3) Para cualquier colección de intervalos no superpuestos $\{I_i\}_{i=1}^n$, si definimos, para $i = 1, \dots, n$, $E_i = \{\text{ocurren exactamente } j_i \text{ sucesos en } I_i\}$, con $j_i \in \mathbb{N}_0$, entonces los eventos E_1, \dots, E_n son independientes.

Notemos que, por ejemplo, $f(h) = h^2$ es $o(h)$, mientras que $f(h) = h$ no lo es. Así, para valores pequeños de h , la condición 1 nos dice que la probabilidad de que exactamente un suceso ocurra en un intervalo de longitud h es aproximadamente proporcional a la longitud del intervalo, siendo igual a λh más algo que es pequeño comparado con h . La condición 2 establece que la probabilidad de que ocurran dos o más sucesos en un intervalo de longitud h es pequeña comparada con h . La condición 3 indica que lo que ocurre en un intervalo no tiene efecto en lo que ocurre en otro intervalo no superpuesto.

Sea $t > 0$ y $N(t)$ la variable aleatoria que cuenta el número de veces que ocurre el suceso o fenómeno bajo estudio en el intervalo de tiempo $[0, t]$. Bajo las condiciones recién planteadas, veamos que $N(t)$ resulta tener una distribución de Poisson de parámetro λt . Para calcular $P(N(t) = k)$, dividamos el intervalo $[0, t]$ en n subintervalos $\{I_i\}_{i=1}^n$ que no se intersecan, cada uno de longitud $\frac{t}{n}$. Esto es, $I_1 = [0, \frac{t}{n}]$ e $I_i = \left(\frac{(i-1)t}{n}, \frac{it}{n}\right]$, $i = 2, \dots, n$. Entonces,

$$\begin{aligned} P(N(t) = k) &= P(k \text{ subintervalos contienen exactamente 1 suceso y los} \\ &\quad \text{restantes } n - k \text{ contienen 0 sucesos}) \\ &\quad + P(N(t) = k \text{ y al menos un subintervalo contiene dos} \\ &\quad \text{o más sucesos}) \\ &= P(A_n) + P(B_n). \end{aligned}$$

Analicemos $P(B_n)$:

$$\begin{aligned} 0 \leq P(B_n) &\leq P\left(\bigcup_{i=1}^n \{I_i \text{ contiene dos o más sucesos}\}\right) \\ &\leq \sum_{i=1}^n P(\{I_i \text{ contiene dos o más sucesos}\}) = no\left(\frac{t}{n}\right) \end{aligned}$$

por la suposición 2. Para cada t fijo, se cumple que $\frac{t}{n} \rightarrow 0$ para $n \rightarrow \infty$. Luego, por la definición de $o(h)$, se cumple que $\lim_{n \rightarrow \infty} o\left(\frac{t}{n}\right) / \left(\frac{t}{n}\right) = 0$ y como $no\left(\frac{t}{n}\right) = t \left[o\left(\frac{t}{n}\right) / \left(\frac{t}{n}\right)\right]$, resulta que $P(B_n) \xrightarrow{n \rightarrow \infty} 0$.

Ahora veamos $P(A_n)$. Dado un intervalo I de longitud h ,

$$1 = P(\text{ningún suceso ocurre en } I) + P(\text{exactamente un suceso ocurre en } I) + P(\text{dos o más sucesos ocurren en } I).$$

De las suposiciones 1 y 2 se cumple que

$$1 = P(\text{ningún suceso ocurre en } I) + [\lambda h + o(h)] + o(h).$$

La suma de funciones $o(h)$ es $o(h)$, pues si $\lim_{h \rightarrow 0} \frac{f(h)}{h} = 0$ y $\lim_{h \rightarrow 0} \frac{g(h)}{h} = 0$, entonces $\lim_{h \rightarrow 0} \frac{f(h)+g(h)}{h} = 0$. Así,

$$P(\text{ningún suceso ocurre en } I) = 1 - \lambda h - o(h).$$

Aplicando la suposición de independencia dada en la condición 3 a los intervalos I_i definidos arriba, de longitud $\frac{t}{n}$, resulta que

$$P(A_n) = \binom{n}{k} \left[\frac{\lambda t}{n} + o\left(\frac{t}{n}\right) \right]^k \left[1 - \frac{\lambda t}{n} - o\left(\frac{t}{n}\right) \right]^{n-k}.$$

Recordando que $\binom{n}{k} = \frac{n(n-1)\cdots(n-(k-1))}{k!}$, resulta que

$$\begin{aligned} \lim_{n \rightarrow \infty} P(A_n) &= \frac{1}{k!} \lim_{n \rightarrow \infty} \left(\prod_{j=0}^{k-1} \left\{ (n-j) \left[\frac{\lambda t}{n} + o\left(\frac{t}{n}\right) \right] \right\} \right) \\ &\quad \times \lim_{n \rightarrow \infty} \left[1 - \frac{\lambda t}{n} - o\left(\frac{t}{n}\right) \right]^n \lim_{n \rightarrow \infty} \left[1 - \frac{\lambda t}{n} - o\left(\frac{t}{n}\right) \right]^{-k}. \end{aligned}$$

Para calcular el primero de los límites del término de la derecha, observemos que, si $0 \leq j \leq k-1$,

$$(n-j) \left[\frac{\lambda t}{n} + o\left(\frac{t}{n}\right) \right] = \lambda t + n o\left(\frac{t}{n}\right) - j \left[\frac{\lambda t}{n} + o\left(\frac{t}{n}\right) \right] \xrightarrow{n \rightarrow \infty} \lambda t,$$

pues $o(t/n) = (t/n) \frac{o(t/n)}{(t/n)} \xrightarrow{n \rightarrow \infty} 0$. Entonces,

$$\lim_{n \rightarrow \infty} \prod_{j=0}^{k-1} \left\{ (n-j) \left[\frac{\lambda t}{n} + o\left(\frac{t}{n}\right) \right] \right\} = (\lambda t)^k.$$

Para el segundo límite, sea $a_n = \lambda t - n o\left(\frac{t}{n}\right)$. Como $\lim_{n \rightarrow \infty} a_n = \lambda t$, entonces

$$\lim_{n \rightarrow \infty} \frac{n}{a_n} = \infty.$$

Usando que

$$\lim_{n \rightarrow \infty} \left(1 + \frac{\delta}{n} \right)^n = e^\delta,$$

para $\delta = -1$ resulta que

$$\lim_{n \rightarrow \infty} \left(1 - \frac{1}{n/a_n} \right)^{n/a_n} = e^{-1}.$$

Luego,

$$\begin{aligned} \lim_{n \rightarrow \infty} \left[1 - \frac{\lambda t}{n} - o\left(\frac{t}{n}\right) \right]^n &= \lim_{n \rightarrow \infty} \left(1 - \frac{a_n}{n} \right)^n = \lim_{n \rightarrow \infty} \left[\left(1 - \frac{1}{\frac{n}{a_n}} \right)^{\frac{n}{a_n}} \right]^{a_n} \\ &= \lim_{n \rightarrow \infty} \exp \left\{ a_n \log \left[\left(1 - \frac{1}{\frac{n}{a_n}} \right)^{\frac{n}{a_n}} \right] \right\} \\ &= \exp \left\{ \left(\lim_{n \rightarrow \infty} a_n \right) \log \left[\lim_{n \rightarrow \infty} \left(1 - \frac{1}{\frac{n}{a_n}} \right)^{\frac{n}{a_n}} \right] \right\} \\ &= \exp(-\lambda t). \end{aligned}$$

Por último,

$$\lim_{n \rightarrow \infty} \left[1 - \frac{\lambda t}{n} - o\left(\frac{t}{n}\right) \right]^{-k} = 1,$$

con lo cual

$$P(A_n) \xrightarrow{n \rightarrow \infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!}.$$

Así, como $P(N(t) = k) = P(A_n) + P(B_n)$, haciendo tender n a ∞ se obtiene

$$P(N(t) = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}, \quad k = 0, 1, \dots,$$

resultando que $N(t) \sim \mathcal{P}(\lambda t)$. La cantidad λ corresponde al número esperado de eventos por unidad de tiempo. \square

Proposición (Caracterización de la distribución geométrica): Sea Y una variable aleatoria discreta con valores posibles $1, 2, \dots$ tal que

$$P(Y > n + m) = P(Y > n)P(Y > m) \text{ para todo } n, m \in \mathbb{N}_0.$$

Entonces, Y tiene distribución geométrica.

Demostración (Hoel y cols., 1971): La propiedad

$$P(Y > n + m) = P(Y > n)P(Y > m)$$

vale para todo $n, m \in \mathbb{N}_0$, en particular para $n = m = 0$. Por lo tanto, $P(Y > 0) = P(Y > 0)^2$ y en consecuencia $P(Y > 0)$ vale 0 o 1. Pero como Y toma valores positivos, necesariamente se cumple que $P(Y > 0) = 1$.

Sea $p = P(Y = 1)$ (con $0 < p < 1$, pues $\sum_{i=1}^{\infty} P(Y = i) = 1$ y $P(Y = i) > 0$ para todo i). Tomando $m = 1$, y repitiendo el razonamiento, se cumple para todo $n \in \mathbb{N}$ que

$$\begin{aligned} P(Y > n) &= P(Y > (n-1) + 1) = P(Y > n-1)P(Y > 1) \\ &= P(Y > n-1)(1-p) = P(Y > n-2)(1-p)^2 \\ &= \dots = P(Y > 0)(1-p)^n = (1-p)^n. \end{aligned}$$

Como $P(Y > n-1) = P(Y \geq n) = P(Y > n) + P(Y = n)$, resulta

$$\begin{aligned} P(Y = n) &= P(Y > n-1) - P(Y > n) \\ &= (1-p)^{n-1} - (1-p)^n \\ &= (1-p)^{n-1}(1 - (1-p)) = (1-p)^{n-1}p. \end{aligned}$$

Es decir, $Y \sim \mathcal{G}(p)$. \square

Proposición (Caracterización de la distribución exponencial): Sea Y una variable aleatoria tal que

$$(5.1) \quad P(Y > a + b) = P(Y > a)P(Y > b) \text{ para todo } a, b \geq 0.$$

Entonces, $P(Y > 0) = 0$ o Y tiene distribución exponencial.

Demostración (Hoel y cols., 1971): Apliquemos (5.1) a $a = b = 0$. Escribiendo $0 = 0 + 0$, resulta $P(Y > 0) = P(Y > 0 + 0) = P(Y > 0)^2$, con lo cual $P(Y > 0)$ vale 0 o 1.

Si $P(Y > 0) = 0$, siempre se cumple (5.1) ya que si $c \geq 0$, el evento $\{Y > c\}$ está contenido en el evento $\{Y > 0\}$ y por lo tanto $0 \leq P(Y > c) \leq P(Y > 0) = 0$, resultando $P(Y > c) = 0$. Así, todos los términos que intervienen en (5.1) son 0. Pero esta propiedad nos dice poco respecto a Y ya que esencialmente Y no toma valores positivos ($P(Y \leq 0) = 1$).

El caso más interesante ocurre cuando $P(Y > 0) = 1$. Veremos que necesariamente Y , por cumplir (5.1), debe tener distribución exponencial.

Sea $G(y) = P(Y > y) = 1 - F(y)$, con F la función de distribución de Y . Luego, por las propiedades que cumple F , G resulta ser una función continua a derecha, no creciente (si $x < y \Rightarrow G(x) \geq G(y)$) y $\lim_{y \rightarrow \infty} G(y) = 1 - \lim_{y \rightarrow \infty} F(y) = 0$. Además, $G(0) = P(Y > 0) = 1$ y

$$G(a + b) = P(Y > a + b) = P(Y > a)P(Y > b) = G(a)G(b)$$

para todo $a, b > 0$. De esto se deduce lo siguiente:

(i) Si $n \in \mathbb{N}$ y $c > 0$, entonces

$$G(nc) = G(\underbrace{c + \dots + c}_{n \text{ veces}}) = G(\underbrace{c + \dots + c}_{n-1 \text{ veces}})G(c) = \dots = \underbrace{G(c) \dots G(c)}_{n \text{ veces}} = G(c)^n.$$

(ii) Si $c > 0$ y $m \in \mathbb{N}$, entonces

$$G(c) = G\left(\frac{c}{m}m\right) = \left(G\left(\frac{c}{m}\right)\right)^m$$

por (i), pues $\frac{c}{m} > 0$ y $m \in \mathbb{N}$.

Notemos que $0 \leq G(y) \leq 1$ por ser una probabilidad. Pero la desigualdad es estricta para $y = 1$ (esto es, $0 < G(1) < 1$), pues

- Si $G(1) = 1$, entonces por (i) tomando $c = 1$ resulta $G(n) = (G(1))^n = 1$ para todo $n \in \mathbb{N}$. Esto es absurdo pues $\lim_{y \rightarrow \infty} G(y) = 0$.
- Si $G(1) = 0$, entonces, por (ii), tomando $c = 1$ resulta $0 = G(1) = (G(\frac{1}{m}))^m$ para todo $m \in \mathbb{N}$. Luego, $G(\frac{1}{m}) = 0$ para todo $m \in \mathbb{N}$. Así, $\{\frac{1}{m}\}_{m \in \mathbb{N}}$ es una sucesión que tiende a 0 por valores mayores a 0 tal que $G(\frac{1}{m})$ tiende a 0. Esto es absurdo pues G es continua por derecha, con lo cual $\lim_{y \rightarrow 0^+} G(y) = G(0) = 1$.

Así, como $0 < G(1) < 1$, podemos escribir $G(1) = e^{-\lambda}$ para algún $\lambda > 0$ (precisamente, $\lambda = -\log(G(1))$). Veamos ahora que $G(y) = e^{-\lambda y}$ para todo $y > 0$. Primero lo veremos para los racionales. Usando ii) para $c = 1$, resulta

$$G\left(\frac{1}{m}\right) = (G(1))^{\frac{1}{m}} = e^{-\lambda \frac{1}{m}}.$$

Además, si $m, n \in \mathbb{N}$, entonces por i) tomando $c = \frac{1}{m}$, resulta

$$G\left(\frac{n}{m}\right) = \left(G\left(\frac{1}{m}\right)\right)^n = \left(e^{-\lambda\frac{1}{m}}\right)^n = e^{-\lambda\frac{n}{m}}.$$

En consecuencia, $G(q) = e^{-\lambda q}$ para todo $q > 0 \in \mathbb{Q}$. Ahora sea $y > 0$. Luego, existe una sucesión $\{q_n\}_{n \in \mathbb{N}}$ de racionales positivos mayores a y tal que $\lim_{n \rightarrow \infty} q_n = y$. Entonces, como G es continua por derecha, se cumple

$$G(y) = \lim_{n \rightarrow \infty} G(q_n) = \lim_{n \rightarrow \infty} e^{-\lambda q_n} = e^{-\lambda y},$$

ya que $g(x) = e^{-\lambda x}$ es una función continua. Así, $G(y) = e^{-\lambda y}$ para todo $y > 0$. Por otro lado, $G(0) = 1$ y $G(y) = 1$ para todo $y < 0$ pues como G es no creciente, $1 \geq G(y) \geq G(0) = 1$. De esta manera,

$$G(y) = \begin{cases} e^{-\lambda y}, & y > 0, \\ 1, & \text{caso contrario,} \end{cases}$$

Finalmente,

$$F(y) = \begin{cases} 1 - e^{-\lambda y}, & y > 0, \\ 0, & \text{caso contrario} \end{cases}$$

con lo cual $Y \sim \text{Exp}(\lambda)$, como queríamos demostrar. \square

Bibliografía

- Agresti, A., Franklin, C., y Klingenberg, B. (2018). *Statistics: The art and science of learning from data* (4th ed.). Harlow: Pearson.
- Ash, R. B. (2008). *Basic probability theory*. Mineola: Dover Publications.
- Bertsekas, D. P., y Tsitsiklis, J. N. (2008). *Introduction to probability* (2nd ed.). Belmont: Athena Scientific.
- Blitzstein, J. K., y Hwang, J. (2015). *Introduction to probability*. Boca Raton: CRC Press.
- Canavos, G. (1995). *Probabilidad y estadística: Aplicaciones y métodos*. México: Mc Graw Hill.
- Devore, J. L. (2005). *Probabilidad y estadística para ingeniería y ciencias* (6ta ed.). México: Thomson.
- Grimmett, G., y Stirzaker, D. (2001). *Probability and random processes* (3rd ed.). Oxford: Oxford University Press.
- Hoel, P. G., Port, S. C., y Stone, C. J. (1971). *Introduction to probability theory*. Boston: Houghton Mifflin Company.
- R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Descargado de <https://www.R-project.org/>
- Ross, S. M. (1999). *Simulación* (2da ed.). México: Prentice Hall.
- Ross, S. M. (2014). *A first course in probability* (9th ed.). Boston: Pearson.

Wackerly, D. D., Mendenhall, W., y Scheaffer, R. L. (2010). *Estadística matemática con aplicaciones* (7ma ed.). México: Cengage Learning.

MARÍA LAURA NORES

Facultad de Matemática, Astronomía, Física y Computación, Universidad Nacional de Córdoba.

(✉) mlnores@famaf.unc.edu.ar

Recibido: 17 de octubre de 2020.

Aceptado: 3 de marzo de 2021.

Publicado en línea: 31 de julio de 2021.
