Clase 8: Valor esperado

Matías Carrasco

13 de agosto de 2019

Resumen Motivados por argumentos intuitivos veremos cómo calcular el valor promedio (esperanza) de una variable aleatoria discreta. Una aplicación interesante de este concepto es el famoso "problema de las muestras sanguíneas". Luego generalizamos esta noción a variables continuas, y tratamos algunos ejemplos.

Índice

Definición informal de valor esperado

Comencemos por un ejemplo simple. Supongamos que estamos al frente de un pequeño emprendimiento en el cual cada semana debemos decidir entre dos opciones:

1. cerramos un negocio seguro que nos provee una ganancia neta de \$15.000;
2. o realizamos una inversión que de salir bien nos aportaría una ganancia neta de \$30.000, pero de salir mal conllevaría una perdida neta de \$15.000. Además, estimamos que la probabilidad de que la inversión sea exitosa es de 75 %.

Imaginemos que tomamos la decisión de invertir en n semanas consecutivas. Denotemos por n_+ el número de veces que la inversión ha resultado exitosa, y $n_- = n - n_+$ el número de veces que dio pérdidas. Entonces, las ganancias totales $G(n)$ (en miles de pesos) en esas n semanas son

$$G(n) = 30n_+ - 15n_-.$$

Las ganancias por semana son

$$g(n) = \frac{G(n)}{n} = 30\left(\frac{n_+}{n}\right) - 15\left(\frac{n_-}{n}\right).$$

¿Qué ocurre a la larga con las ganancias por semana? Sabemos que las frecuencias relativas n_+/n y n_-/n convergen, cuando n tiende a infinito, a las respectivas probabilidades de que la inversión sea exitosa o fracase[1]. Entonces

$$\begin{aligned}\lim_{n\to\infty} g(n) &= 30 \lim_{n\to\infty} \frac{n_+}{n} - 15 \lim_{n\to\infty} \frac{n_-}{n} \\ &= 30 \cdot (\text{prob. de éxito}) - 15 \cdot (\text{prob. de fracaso}) \\ &= 30 \cdot 0.75 - 15 \cdot 0.25 = 18.75.\end{aligned}$$

[1] Al menos con probabilidad alta. Recordar que la probabilidad de un evento da el porcentaje de tiempo que se espera que ocurra, cuando el proceso básico se realiza una y otra vez, de forma independiente y en las mismas condiciones.

Es decir, a medida que n crece, las ganancias por semana se aproximan más y más al valor $18.750.

Si hubiéramos optado por la opción segura, las ganancias por semana serían iguales a $g(n) = 15$. Como las ganancias por semana son mayores para la opción 2 que para la opción 1, es mejor arriesgar invirtiendo el dinero, siempre y cuando seamos capaces de invertir durante una cantidad grande de semanas.

Si pensamos a las ganancias semanales como una variable aleatoria G, que toma los valores 30 y -15, con probabilidades 0.75 y 0.25 respectivamente, entonces la cantidad 18.75 se llama *el valor esperado* de G. Esto lo escribimos $E(G)$.

La definición formal para variables discretas

La misma idea nos sirve como motivación para definir el valor esperado de una variable discreta en general. Supongamos que X es una variable discreta cuyo recorrido es $R_X = \{x_1, x_2, \ldots\}$. Imaginemos que realizamos el experimento n veces y para cada una de estas registramos el valor de X. Llamemos a estos valores por $y_1, y_2, \ldots, y_n$. Cada uno de los y_i puede ser igual a cualquiera de los valores posibles de X (los valores del recorrido de X).

El promedio de las n realizaciones de X es

$$\text{Promedio}(y_1, \ldots, y_n) = \frac{y_1 + \cdots + y_n}{n}.$$

Podemos reordenar los valores $y_1, \ldots, y_n$ y agruparlos de acuerdo a su valor, de modo que la suma

$$y_1 + \cdots + y_n = n_1 x_1 + n_2 x_2 + \cdots,$$

en donde n_j es el número de veces que ha ocurrido el valor x_j. En símbolos $n_j = \left|\{i : y_i = x_j\}\right|$. Con esta forma de escribir la suma, tenemos que

$$\text{Promedio}(y_1, \ldots, y_n) = x_1 \left(\frac{n_1}{n}\right) + x_2 \left(\frac{n_2}{n}\right) + \cdots.$$

Al realizar más veces el experimento, y hacer n tender a infinito, las frecuencias relativas convergen a

$$\frac{n_j}{n} \to P\left(X = x_j\right).$$

El valor "por ensayo" de X, para n tendiendo a infinito, es $E(X) = \sum_{j=1}^{\infty} x_j P\left(X = x_j\right)$.

Definición de valor esperado (discretas)

Sea X una variable aleatoria discreta cuyo recorrido es R_X. Definimos el valor esperado de X (o la esperanza de X) como

$$E(X) = \sum_{x \in R_X} x \cdot P(X = x).$$

Ejemplo 1 Los primeros en pensar en valor esperado fueron los matemáticos franceses Pascal y Fermat en una vasta correspondencia que iniciaron en 1654. Un noble llamado Chevalier de Méré le propuso a Pascal el siguiente problema:

> Un jugador ha apostado en sacar un 6 en 8 lanzamientos de un dado. El monto ha sido establecido, y se han realizado 3 lanzamientos sin la aparición de un 6. ¿Qué proporción del monto apostado sería justo darle al jugador para que renuncie al cuarto lanzamiento (solo el cuarto)?

Digamos que la apuesta es a. Llamemos X a la ganancia del jugador en la apuesta original, e Y la ganancia si renuncia al cuarto lanzamiento.

Como quedan 5 lanzamientos, la f.p.p. de X es

Valor x	a	0
f.p.p. $p(x)$	$1 - \left(\frac{5}{6}\right)^5$	$\left(\frac{5}{6}\right)^5$

Entonces, el valor esperado de X es $E(X) = \left(1 - \left(\frac{5}{6}\right)^5\right) a$.

Si el jugador renuncia a su cuarto lanzamiento, obtendrá con seguridad una fracción f de la apuesta a. El resto del monto, $(1-f)a$, lo obtendrá con probabilidad $1 - \left(\frac{5}{6}\right)^4$. Entonces, el valor esperado de Y es

$$E(Y) = fa + \left(1 - \left(\frac{5}{6}\right)^4\right)(1-f)a.$$

Para Pascal y Fermat, la proporción justa f es aquella que mantiene las expectativas de ganancia del jugador. Esto se traduce en $E(Y) = E(X)$.

Esta igualdad se traduce en

$$(1-f)\left(\frac{5}{6}\right)^4 = \left(\frac{5}{6}\right)^5,$$

de donde $f = 1/6$. Lo justo es proponerle $1/6$ de la apuesta para que renuncie al cuarto lanzamiento. ■

Para pensar: ¿qué proporción le ofrecerías para renunciar al quinto lanzamiento, si en lugar de 3 han transcurrido 4 lanzamientos sin aparecer un 6?

Variables de Bernoulli

Son los bloques fundamentales a partir de los cuales podemos construir una gran variedad de variables discretas. Las variables Bernoulli indican la ocurrencia o no de un determinado evento de interés. En general se usa la palabra "éxito" para indicar que el evento ocurre, y "fracaso" para indicar que no ocurre.

Una variable X tiene distribución de Bernoulli si

$$X = \begin{cases} 1 & \text{si éxito;} \\ 0 & \text{si fracaso.} \end{cases}$$

Para conocer la distribución de la variable X, basta determinar el parámetro p, que representa la probabilidad de éxito, esto es $P(X=1) = p$. Cuando queremos escribir de forma compacta que X tiene distribución Bernoulli de parámetro p ponemos $X \sim \text{Ber}(p)$.

La esperanza de una variable Bernoulli es

$$E(X) = 1 \cdot P(X=1) + 0 \cdot P(X=0) = p.$$

Esta simple fórmula es de mucha ayuda, ya que en una gran variedad de situaciones podemos descomponer una variable como suma de Bernoulli. El siguiente es un típico ejemplo de este uso.

Ejemplo 2 En un grupo de n personas distintas. ¿Cuántas coincidencias de cumpleaños esperamos ver?

Imaginemos a las personas numeradas del 1 al n. Para cada par de personas $\{i, j\}$, consideremos la variable X_{ij} que vale 1 si i y j cumplen el mismo día, y 0 si no. Claramente X_{ij} es una variable de Bernoulli, pues toma solamente los valores 0 y 1. Aquí éxito corresponde a que i y j cumplan el mismo día.

El parámetro de X_{ij} corresponde a la probabilidad de que valga 1:

$$\begin{aligned} p &= P\left(X_{ij} = 1\right) \\ &= P\left(i \text{ y } j \text{ cumplen en el mismo día}\right) = \frac{1}{365} \end{aligned}$$

Es decir, $X_{ij} \sim \text{Ber}(1/365)$.

Llamemos X a la suma de las X_{ij} sobre todos los pares posibles. Esto es

$$X = \sum_{\{i,j\}} X_{ij}.$$

¿Qué representa X? La variable X cuenta cuántas coincidencias de cumpleaños hay en el grupo de n personas. Así que la probabilidad de que al menos dos personas cumplan el mismo día se puede escribir $P(X \geq 1)$.

Lo bueno de este enfoque es que es fácil pasar de dos personas a tres, cuatro, etc. Simplemente, en lugar de considerar las variables X_{ij} sobre los pares, hay que considerar X_{ijk} sobre las tripletas, etc.

Aceptemos por el momento que la esperanza de una suma es la suma de las esperanzas[2]. El valor esperado de X es entonces

[2] Es una propiedad sumamente importante que demostraremos la clase que viene

$$E(X) = \sum_{\{i,j\}} E(X_{ij}) = \sum_{\{i,j\}} \frac{1}{365} = \binom{n}{2}\frac{1}{365} = \frac{n(n-1)}{730}.$$

Notar que $E(X) \geq 1$ si $n = 28$. Esto sugiere que con 28 personas es altamente probable que haya al menos una coincidencia. ■

Arquímides y el valor esperado

Todos saben que el estudio del centro de masa de un objeto fue la devoción de Arquímedes, el matemático de la Grecia antigua. Seguramente conozcan su famosa frase "Dadme un punto de apoyo, y moveré al mudo". En su descubrimiento de la "ley de la palanca", Arquímedes demuestra cómo encontrar el punto de equilibrio para configuraciones de objetos similares a los subibaja que tanto disfrutamos en nuestra niñez.

Comencemos por el caso más simple de todos: dos bloques de 0.5 kg cada uno apoyados sobre un tablón, que se balancea sobre un pie de apoyo como un subibaja. ¿En dónde deberíamos colocar el pie para que el tablón quede en equilibrio? Obviamente, por la simetría del problema, en el punto medio entre los dos bloques.

Otra forma de llegar a esta conclusión es la siguiente: si tenemos un solo bloque de 1 kg, es obvio que el punto de apoyo debe situarse justo debajo del bloque. Si ahora partimos el bloque en dos mitades iguales, y las desplazamos igual distancia hacia derecha y izquierda, podemos dejar el pie siempre en el mismo lugar y el conjunto quedará en equilibrio todo el tiempo. Ver la Figura 1. Esto es porque no hemos cambiado el centro de masa del conjunto de bloques.

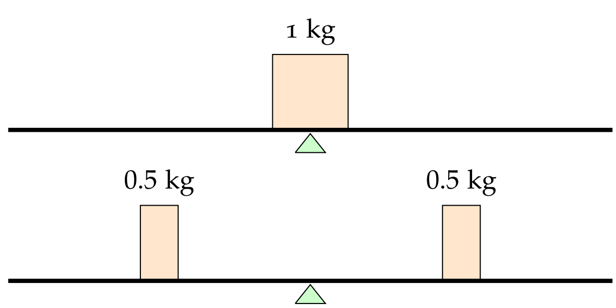

Figura 1: Dos bloques de 0.5 kg se mantienen en equilibrio colocando el pie en el punto medio entre ellos.

Usaremos este principio, de que dos configuraciones de bloques con el mismo centro de masa producen el mismo efecto sobre el tablón, para calcular el centro de masa de una configuración cualquiera.

¿Qué pasa si en lugar de dividir el bloque original de 1 kg en dos mitades iguales, lo dividimos en un bloque de p kg y otro de q kg, con $p + q = 1$?

Por ejemplo, $p = 1/3$ y $q = 2/3$. Intuitivamente, el pie lo debemos colocar dos veces más cerca del bloque más pesado que del bloque más liviano. Demostremos que ésto es así.

Supongamos que los bloques están a distancia 1. Dividimos la distancia entre los dos bloques en tres partes iguales, de forma que el pie está situado a distancia $1/3$ del bloque pesado, como se muestra en la Figura 2.

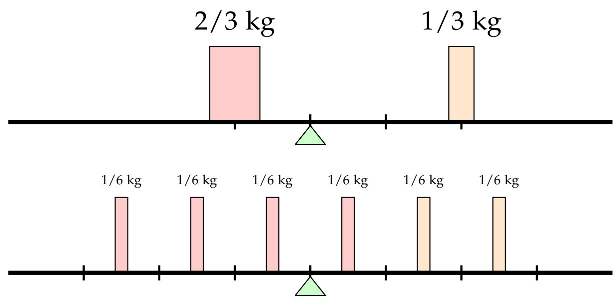

Figura 2: Dos bloques, uno de $1/3$ kg y el otro de $2/3$ kg, se mantienen en equilibrio colocando el pie a una distancia $1/3$ del bloque más pesado.

Marquemos dos segmentos de longitud $1/3$ hacia la izquierda del bloque pesado y uno hacia la derecha del bloque liviano. Si cambiamos los dos bloques por seis bloques que pesan $1/6$ kg, colocados sobre los puntos medios de los segmentos marcados, obtenemos una configuración equivalente, pues no hemos cambiado los centros de masa.

Como la nueva configuración de bloques es simétrica, es claro que el pie de apoyo debe ir en el centro, lo cual implica que la configuración original estaba en equilibrio.

El mismo argumento se puede hacer para cualquier par de bloques cuyos pesos seas racionales. Luego, usando un pasaje al límite se puede extender el resultado al caso de pesos irracionales.

La conclusión es que si colocamos dos bloques que pesan p y q kilogramos, con el bloque que pesa p en la posición 1 y el bloque que pesa q en la posición 0, entonces el pie de apoyo debe colocarse en la posición p. Ver la Figura 3.

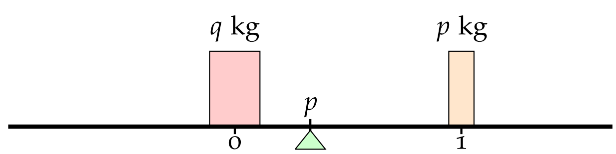

Figura 3: Dos bloques, uno de p kg y el otro de q kg, se mantienen en equilibrio colocando el pie a una distancia p del origen.

Notar que los bloques de la Figura 3 representan la distribución de una variable Bernoulli de parámetro p. Como la esperanza de una tal variable es igual a p, hemos probado que la posición del pie de apoyo coincide con el valor esperado. Esto es completamente general.

Podemos representar la distribución de cualquier variable discreta X usando bloques y un tablón. Primero, marcamos un origen cualquiera en el tablón desde el cual medir distancias. Si X toma los valores

$x_1, x_2, \ldots$ con probabilidades $p_1, p_2, \ldots$, colocamos un bloque que pesa p_i kg en la posición x_i. Notar que el peso total de los bloques es 1 kg pues las probabilidades suman 1. Entonces:

> El valor esperado $E(X)$ indica en dónde debe colocarse el pie de apoyo para que el conjunto de bloques se mantenga en equilibrio.

Dicho de otro modo, el valor esperado es el centro de masa de la distribución.

El problema de las pruebas sanguíneas

Un gran número N de personas se ve sometido a una prueba sanguínea, que se puede administrar en dos formas:

(i) Se prueba a cada persona por separado. En este caso se requieren N análisis.

(ii) Se mezclan las muestras de sangre de k personas para examinarlas juntas. Si la prueba resulta negativa, esta prueba será suficiente para las k personas. Si la prueba es positiva, cada una de las k personas debe analizarse separadamente, y en total, se requieren $k+1$ pruebas para las k personas.

Supongamos que la probabilidad p de que el análisis resulte positivo es la misma para todas las personas, y que los resultados de las pruebas son independientes.

¿Cuál es la probabilidad de que el análisis de una muestra mezclada de k personas resulte positivo? Como las personas resultan positivas con la misma probabilidad p y son independientes, la probabilidad de que ninguna de las personas del grupo sea positiva es $(1-p)^k$. Así que la probabilidad de que el grupo resulte positivo, o sea de que alguna de ellas sea positiva, es $q = 1-(1-p)^k$.

¿Cuál es el valor esperado del número X de pruebas necesarias en el programa (ii)? Aunque a primera vista parezca difícil calcular el valor esperado de X, usaremos un truco muy común que consiste en escribir la variable como suma de variables más simples. La variable X se puede escribir como

$$X = \sum_{i=1}^{N/k} X_i$$

en donde X_i es el número de pruebas necesarias para el i-ésimo grupo. Esta puede tomar dos valores

$$X_i = \begin{cases} 1 & \text{si el grupo da negativo} \\ k+1 & \text{si el grupo da positivo.} \end{cases}$$

Las probabilidades son respectivamente $1-q$ y q.

Ahora podemos usar la propiedad de linealidad del valor esperado. La esperanza de cada X_i es igual a

$$E(X_i) = 1 \cdot (1-q) + (k+1) \cdot q = 1 + kq.$$

Entonces, el valor esperado de X es

$$E(X) = \frac{N}{k}(1+kq) = N\left(\frac{1}{k} + q\right).$$

Supongamos ahora que se trata de una prueba para una enfermedad rara, para la cual p es muy chico. Entonces, podemos aproximar $(1-p)^k$ por $1-kp$ ya que los otros términos contienen potencias mayores de p. Con esta aproximación el valor esperado nos queda

$$E(X) \approx N\left(\frac{1}{k} + kp\right).$$

¿Cuál es el valor de k que minimiza el valor esperado? Si obviamos el hecho de que k debe ser entero y consideramos la función

$$f(x) = \frac{1}{x} + xp,$$

usando las herramientas de cálculo vemos que el mínimo se da en $x = 1/\sqrt{p}$.

Es decir, si p es chico, el método de mezclar las muestras de grupos de tamaño $k \approx 1/\sqrt{p}$ hace que el valor esperado de pruebas se reduzca a $E(X) \approx 2\sqrt{p}N$.

Valor esperado de variables continuas

¿Cómo podemos definir la esperanza de una variable continua X? La idea es muy simple. Imaginemos que dividimos la recta real en intervalos de longitud Δx_k centrados en puntos x_k. Llamemos X_0 a la variable discreta que vale x_k cuando X cae en el intervalo centrado en x_k. Notar que la probabilidad de que X_0 valga x_k es aproximadamente igual a $p(x_k)\Delta x_k$.

La esperanza de X_0 es por definición

$$E(X_0) = \sum_k x_k P(X_0 = x_k) = \sum_k x_k p(x_k)\Delta x_k.$$

Si los intervalos Δx_k son todos pequeños, esta suma es aproximadamente igual a la integral de $xp(x)$.

Esperanza de una variable continua

La esperanza de una variable continua X con densidad $p(x)$ se define como

$$E(X) = \int_{-\infty}^{+\infty} xp(x)dx.$$

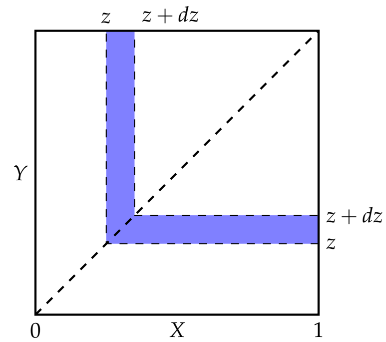

Figura 4: El evento $\{z \le Z \le z+dz\}$.

Ejemplo 3 Dos puntos se eligen al azar de manera uniforme e independiente en un segmento de longitud unidad, y dividen al segmento en 3 piezas. ¿Cuál es la longitud promedio del segmento izquierdo?

Llamemos X e Y a las coordenadas de los dos puntos en el intervalo $[0,1]$. El segmento izquierdo es el que va desde 0 hasta $Z = \min\{X,Y\}$, y su longitud es precisamente Z. Debemos hallar la densidad de Z.

En la Figura 4 se muestra el evento $\{z \le Z \le z+dz\}$. Su área es

$$(1-z)^2 - (1-z-dz)^2 = 2(1-z)dz - dz^2.$$

Luego, la densidad de Z está dada por (para $0 < z < 1$)

$$p(z) = \lim_{dz\to 0} \frac{P(z \le Z \le z+dz)}{dz} = \lim_{dz\to 0} 2(1-z) - dz = 2(1-z).$$

Entonces, la esperanza de Z es

$$E(Z) = \int_0^1 zp(z)dz = 2\int_0^1 z(1-z)dz = \frac{1}{3}.$$

Notar que $1/3$ corresponde a dividir el segmento en 3 partes iguales. ¿Se te ocurre algún argumento de simetría que permita llegar directamente al $1/3$ sin hacer cálculos? ■

Ejemplo 4 El ejemplo anterior nos enseña algo interesante. Cuando tenemos dos puntos aleatorios en un segmento es natural llamar A al más chico y B al más grande. Eso parece simplificar la discusión de casos para, por ejemplo, calcular una probabilidad. Pero eso no es correcto, y el problema es que si llamamos al más chico A, entonces A no tiene distribución uniforme en el segmento.

Usando la notación del ejemplo anterior, $A = \min\{X,Y\}$ y $B = \max\{X,Y\}$. Como vimos, el valor esperado de A es $E(A) = 1/3$. ¿Cuál es el valor esperado de X?

La densidad de X es uniforme en $[0,1]$, por lo que $p(x) = 1$ si $0 < x < 1$. Entonces

$$E(X) = \int_0^1 xp(x)dx = \int_0^1 xdx = \frac{1}{2}.$$

En particular, $E(X) \neq E(A)$, por lo que A y X tienen distribuciones diferentes. ■

Una fórmula para variables positivas

Existe un truco para variables enteras positivas que simplifica a veces las cuentas.

Esperanza de variables enteras positivas

Sea X una variable discreta que toma valores enteros mayores o iguales a cero. Entonces

$$E(X) = \sum_{k=0}^{\infty} P(X > k). \tag{1}$$

Demostración. Por definición tenemos que

$$E(X) = \sum_{k=0}^{\infty} kP(X = k) = \sum_{k=1}^{\infty} kP(X = k),$$

pues el primer término ($k = 0$) es cero.

Podemos poner estos términos en un arreglo triangular

$$\begin{aligned} E(X) = P(X = 1) & \\ + P(X = 2) + P(X = 2) & \\ + P(X = 3) + P(X = 3) + P(X = 3) + \cdots & \end{aligned}$$

Si sumamos las columnas, obtenemos

$$P(X > 0) + P(X > 1) + P(X > 2) + \cdots,$$

que es exactamente lo que queríamos probar. □

Ejemplo 5 Sea X el número de lanzamientos necesarios para que una moneda salga cara. Supondremos que la probabilidad de cara es p.

La probabilidad de $\{X > k\}$ es $(1 - p)^k$, pues una forma equivalente de describir este evento es que los primeros k lanzamientos sean cruz.

Usando la fórmula (1) obtenemos

$$E(X) = \sum_{k=0}^{\infty} (1 - p)^k = \frac{1}{1 - (1 - p)} = \frac{1}{p}.$$

Por ejemplo, si la moneda es justa, se espera lanzar en promedio 2 veces la moneda para que salga cara. ■

Ejemplo 6 Los habitantes de una isla remota planean sus familias teniendo hijos hasta que nazca la primera niña. Vamos a asumir que la probabilidad de tener una niña es 0.5, que los nacimientos son independientes, y que no hay nacimientos múltiples.

¿Cuál es el ratio de niños y niñas en la isla?

Para una familia dada, el número de hijos X puede modelarse como la cantidad de lanzamientos necesarios para obtener cara, cuando se lanza una moneda con probabilidad de cara igual a 0.5. Esta familia tiene entonces $X - 1$ niños y 1 niña. Así que el ratio de niños sobre niñas, para esta familia, es $X - 1$.

El ratio en la isla corresponde a $E(X-1) = 1$. Así que por más raro que parezca, la cantidad de niñas es aproximadamente igual a la de niños. ■

Esta fórmula se puede generalizar a variables discretas positivas que no son necesariamente enteras. La clave está en observar que $P(X > k) = 1 - F(k)$ en donde F es la f.d.a. de X.

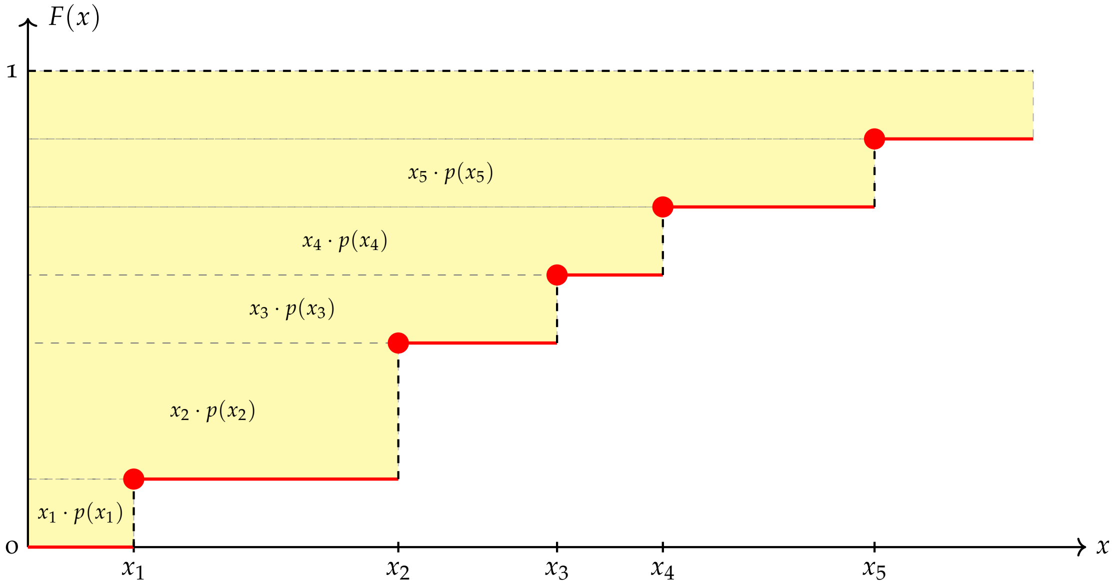

Figura 5: Podemos calcular el valor esperado de una variable positiva X como el área comprendida entre la fda $F(x)$ y 1.

En la Figura 5 están representados los términos $x_i \cdot P(X = x_i)$ como áreas de rectángulos. Al sumar todos estos valores, el resultado es el área total por encima del gráfico de F. Esta área es igual a la integral de $1 - F(x)$ en el intervalo $[0, \infty)$.

Esperanza de una variable positiva

Sea X una variable aleatoria mayor o igual a cero. Entonces

$$E(X) = \int_0^\infty (1 - F(x))dx. \tag{2}$$

Esta fórmula nos será muy útil más adelante. La hemos probado (en la Figura 5) para variables discretas, pero también es válida para variables continuas, y la prueba es aun más simple:

$$\begin{aligned}\int_0^\infty (1-F(x))dx &= \int_0^\infty P(X>x)dx = \int_0^\infty \int_x^\infty p(u)dudx \\ &= \int_0^\infty \int_0^u p(u)dxdu = \int_0^\infty p(u)\int_0^u dxdu \\ &= \int_0^\infty up(u)du = E\left(X\right).\end{aligned}$$

Ejemplo 7 Sea X una variable aleatoria con densidad $p(x) = e^{-x}$ para $x > 0$. ¿Cuál es su valor esperado?

Si aplicamos directamente la definición, debemos calcular la integral

$$E\left(X\right) = \int_0^\infty xp(x)dx = \int_0^\infty xe^{-x}dx.$$

Esta integral se puede resolver por integración por partes. No es difícil ver que la integral vale 1, por lo que $E\left(X\right) = 1$.

Sin embargo, podemos aplicar la fórmula para variables positivas. La fda de X es

$$F(x) = \int_0^x e^{-u}du = 1 - e^{-x}.$$

Luego

$$E\left(X\right) = \int_0^\infty (1-F(x))dx = \int_0^\infty e^{-x}dx = 1.$$

En muchas situaciones similares a esta, la fórmula de variables positivas es más fácil de integrar. ■