Clase 9: Cambio de variable

Matías Carrasco

16 de agosto de 2019

Resumen Estudiaremos cómo calcular la esperanza de una variable que se escribe como función de una o dos variables.

Índice

Esperanza de una función de variables discretas

En general, el procedimiento para calcular el valor esperado de una variable consiste en, primero hallar su f.p.p., y luego aplicar la definición. Sin embargo, cuando la variable de interés se escribe como función de otras, esto puede ser un poco engorroso.

Ejemplo 1 Se apuesta en el lanzamiento de dos dados de la siguiente manera: $Z = XY - 10$, en donde X e Y son el resultado de cada uno de los dados. ¿Cuál es la ganancia esperada?

Como cada resultado tiene probabilidad $1/36$, la f.p.p. de Z la podemos obtener usando la Figura 1, y esta es

-9	-8	-7	-6	-5	-4	-2	-1	0	2	5	6	8	10	14	15	20	26
1	2	2	3	2	4	2	1	2	4	2	1	2	2	2	1	2	1

Valores de Y						
6	-4	2	8	14	20	26
5	-5	0	5	10	15	20
4	-6	-2	2	6	10	14
3	-7	-4	-1	2	5	8
2	-8	-6	-4	-2	0	2
1	-9	-8	-7	-6	-5	-4
	1	2	3	4	5	6
	Valores de X					

Figura 1: Valores posibles de Z según los diferentes valores de X e Y.

Para ser rigurosos, la tabla anterior muestra los diferentes casos, pero para obtener la fpp basta dividir el número dado en la tabla por 36. De aquí resulta entonces $E(Z) = 81/36 = 2.25$. ■

Lo que hicimos en el ejemplo anterior se puede generalizar ampliamente. Si denotamos $g : \mathbb{R}^2 \to \mathbb{R}$ la función $g(x,y) = xy - 10$, entonces $Z = g(X,Y)$.

Esperanza de una función de variables

Sean X e Y dos variables aleatorias discretas, y $g : \mathbb{R}^2 \to \mathbb{R}$ una función. Entonces

$$E(g(X,Y)) = \sum_{x \in R_X, y \in R_Y} g(x,y)p(x,y), \tag{1}$$

en donde $p(x,y)$ es la fpp conjunta de X e Y.

Demostración. El recorrido de $Z = g(X,Y)$ es

$$R_Z = \{g(x,y) : x \in R_X, y \in R_Y\},$$

que es numerable, por lo que $g(X,Y)$ también es discreta.

Notar que estos valores de $g(x,y)$ pueden repetirse, ya que $g(x,y)$ puede ser igual a $g(x',y')$ aunque $x \neq x'$ e $y \neq y'$. De hecho esto pasó en casi todos los casos del Ejemplo 1. Sin embargo, para cada valor posible z de $g(x,y)$, el evento $\{g(X,Y) = z\}$ se descompone como unión disjunta

$$\{g(X,Y) = z\} = \bigcup_{g(x,y)=z} \{X = x, Y = y\}.$$

La unión es en todas las formas distintas de escribir z como $g(x,y)$ para algún x y algún y.

Al tomar probabilidades, obtenemos

$$P\left(g(X,Y) = z\right) = \sum_{g(x,y)=z} P\left(X = x, Y = y\right).$$

Nos resta ahora sumar en z:

$$\begin{aligned} E\left(g(X,Y)\right) &= \sum_{z \in R_Z} z P\left(g(X,Y) = z\right) \\ &= \sum_{z \in R_Z} z \sum_{g(x,y)=z} P\left(X = x, Y = y\right) \\ &= \sum_{z \in R_Z} \sum_{g(x,y)=z} z P\left(X = x, Y = y\right) \\ &= \sum_{z \in R_Z} \sum_{g(x,y)=z} g(x,y) P\left(X = x, Y = y\right) \end{aligned}$$

Aquí viene un punto ligeramente sutil de la demostración. Este consiste en notar que sumar en aquellos x e y con $g(x,y) = z$, y luego sumar en todos los valores posibles de z, es lo mismo que sumar en todos los valores posibles de x e y. Ver la Figura 2.

De aquí resulta que la última suma anterior es igual a

$$\sum_{x \in R_X, y \in R_Y} g(x,y) P\left(X = x, Y = y\right) = \sum_{x \in R_X, y \in R_Y} g(x,y) p(x,y),$$

que es lo que queríamos demostrar. □

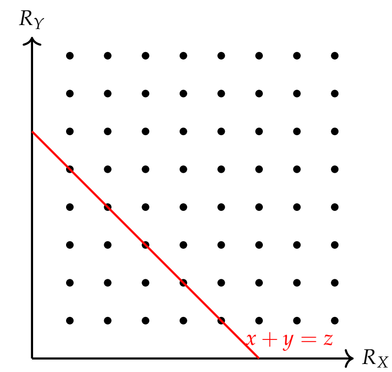


Figura 2: Esto se puede ver mejor con un dibujo, como el que se muestra en esta figura para el caso especial en que $g(x,y) = x + y$. En el mismo vemos la diagonal roja que corresponde a todos los valores de x e y que suman un cierto valor de z. Claramente, al variar z, las diagonales cubren todo el cuadrante. El cuadrante corresponde a todos los pares posibles de x e y.

Cambio de variable

Nuestro objetivo ahora es probar una fórmula análoga a la (1) para el caso de variables continuas. Para esto debemos encontrar la distribución de alguna función de X e Y, digamos $Y = g(X,Y)$, a partir de la distribución conjunta de X e Y. Comenzaremos por el caso de una variable, que es más simple. Luego veremos el caso de dos variables.

Supongamos que X tiene densidad $p_X(x)$. Entonces mientras la función $y = g(x)$ tenga derivada dy/dx que no se anula en ningún intervalo del rango de X, la variable aleatoria $Y = g(X)$ tiene una densidad $p_Y(y)$ que puede calcularse en términos de p_X y la derivada dy/dx. Cómo hacer este cálculo es el tema de esta sección.

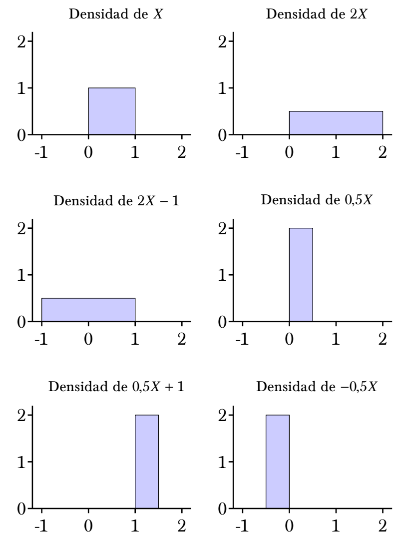


Figura 3: Cambio lineal de variable para densidades uniformes. Los gráficos muestran las densidades de $Y = aX + b$ para varios a y b, donde X tiene distribución uniforme en $(0, 1)$. Observese cómo si $a > 1$ el rango se expande y la densidad disminuye. Y si $0 < a < 1$, el rango se contrae y la densidad aumenta. Al sumar $b > 0$ se traslada hacia la derecha, y al sumar $b < 0$ se traslada hacia la izquierda.

Ejemplo 2 — Cambio de variable lineal. Para ver por qué entra la derivada, observemos primero qué sucede si realizamos un cambio lineal de variable. Para una función lineal $y = ax + b$, la derivada es la constante $dy/dx = a$. La función expande o contrae la longitud de cada intervalo por el mismo factor de $|a|$.

Supongamos que X tiene distribución uniforme en $(0, 1)$, con densidad

$$p_X(x) = \begin{cases} 1 & \text{si } 0 < x < 1; \\ 0 & \text{si no.} \end{cases}$$

Entonces, para $a > 0$ vemos que $Y = aX + b$ tiene distribución uniforme en $(b, b + a)$ con densidad

$$p_Y(x) = \begin{cases} 1/a & \text{si } b < x < b + a; \\ 0 & \text{si no.} \end{cases}$$

De forma similar, si $a < 0$, entonces $Y = aX + b$ tiene distribución uniforme en $(b + a, b)$ con densidad

$$p_Y(x) = \begin{cases} 1/|a| & \text{si } b + a < x < b; \\ 0 & \text{si no.} \end{cases}$$

Se puede pensar que la densidad de $Y = aX + b$ en y es la densidad de X en el punto correspondiente $x = (y - a)/(b - a)$. Pero esto debe dividirse por $|a|$, porque la densidad de probabilidad da la probabilidad por unidad de longitud, y la transformación de x a $ax + b$ multiplica la longitud por el factor $|a|$. Ver la Figura 3.

La densidad de $Y = aX + b$ es entonces

$$p_Y(y) = \frac{1}{|a|} p_X\left(\frac{y - b}{a}\right)$$

Esta fórmula es completamente general. ■

Cambio de variable diferenciable inyectivo

Sea X una variable aleatoria con densidad $p_X(x)$ en el intervalo (a, b). Sea $Y = g(X)$, donde g es estrictamente creciente o estrictamente decreciente en (a, b). Por ejemplo, X podría tener distribución

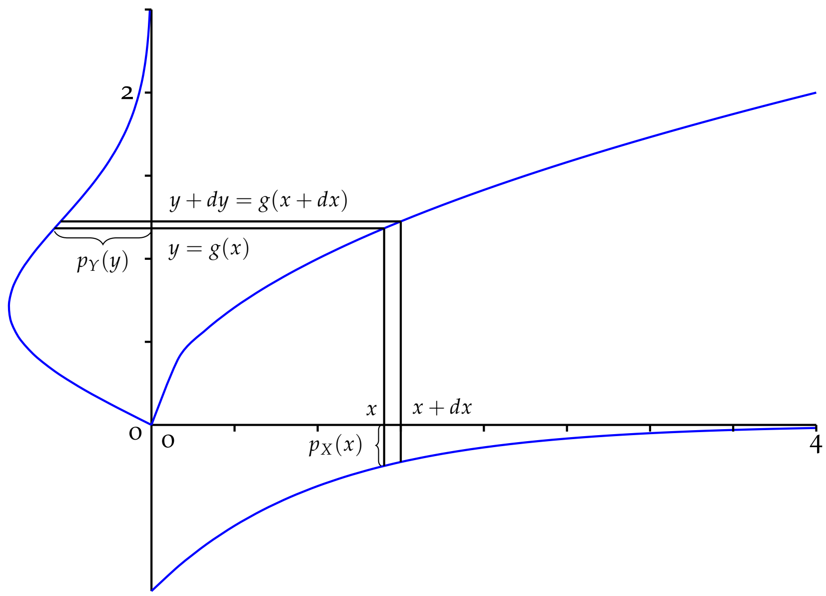


Figura 4: Cambio de variable para densidades. El diagrama muestra el gráfico de $y = g(x)$ para la función creciente $g(x) = \sqrt{x}$, $x > 0$. La densidad $p_X(x)$ se grafica invertida abajo del eje de las x. La densidad $p_Y(y)$ se grafica al costado del eje de las y.

exponencial en $(0, +\infty)$ e Y podría ser X^2, $\sqrt{X}$ o $1/X$. El rango de Y es entonces un intervalo de extremos $g(a)$ y $g(b)$.

El objetivo ahora es calcular la función de densidad $p_Y(y)$ para y en el rango de Y. Para un intervalo infinitesimal dy alrededor de y, el evento $\{Y \in dy\}$ es idéntico al evento $\{X \in dx\}$, donde dx es un intervalo infinitesimal alrededor del único x tal que $y = g(x)$. Ver la Figura 4, donde cada una de las dos áreas sombreadas representa la probabilidad del mismo evento

$$P(Y \in dy) = P(X \in dx) \text{ con } y = g(x).$$

Esta igualdad se traduce en términos de densidades en

$$p_Y(y)dy = p_X(x)dx,$$

y por lo tanto

$$p_Y(y) = p_X(x)\frac{dx}{dy} = \frac{1}{dy/dx}p_X(x) \text{ con } y = g(x).$$

El caso de una función decreciente g es similar, excepto que la derivada dy/dx ahora tiene signo negativo. Este signo debe ser ignorado porque es solo la magnitud de la relación de longitudes de intervalos pequeños lo que es relevante.

Cambio de variable inyectivo

Sea X una variable aleatoria con densidad $p_X(x)$ en el intervalo (a,b), e $Y = g(X)$ con g es creciente o decreciente. Entonces Y toma valores entre $g(a)$ y $g(b)$, con densidad

$$p_Y(y) = \frac{1}{|dy/dx|} p_X(x) \text{ con } y = g(x).$$

La ecuación $y = g(x)$ se debe resolver para x en términos de y, y este valor de x sustituirse en $p_X(x)$ y dy/dx. Esto dará una expresión para $p_Y(y)$ enteramente en términos de y.

Ejemplo 3 Sea X con densidad $p_X(x) = e^{-x}$ para $x \geq 0$. Calculemos la densidad de $Y = \sqrt{X}$.

Para empezar, el recorrido de Y es también el intervalo $[0, +\infty)$. La función $y = g(x) = \sqrt{x}$ es biyectiva en $[0, +\infty)$, y podemos resolver x en función de y pues $x = y^2$. La derivada es $dy/dx = 1/2\sqrt{x}$.

Luego, si sustituimos

$$\frac{1}{|dy/dx|} p_X(x) = \frac{1}{1/2\sqrt{x}} e^{-x}$$

de donde deducimos que $p_Y(y) = 2ye^{-y^2}$ para $y \geq 0$. ■

Cambio de variable en el caso general para $y = g(x)$

Supongamos que la función $y = g(x)$ tiene una derivada que es cero solo en un número finito de puntos. Ahora algunos valores de y pueden provenir de más de un valor de x. Consideremos $Y = g(X)$ para una variable aleatoria X. Como se muestra en la Figura 5, Y estará en un intervalo infinitesimal dy cerca de y cuando X está en uno de los posibles intervalos infinitesimales dx cerca de x tal que $g(x) = y$.

Entonces

$$P(Y \in dy) = \sum_{x:g(x)=y} P(X \in dx),$$

de donde deducimos:

Fórmula general de cambio de variable para $y = g(x)$

$$p_Y(y) = \sum_{x:g(x)=y} \frac{1}{|dy/dx|} p_X(x).$$

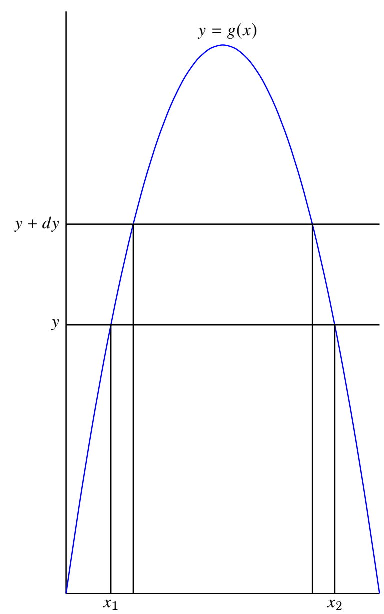


Figura 5: Varios valores de x pueden dar el mismo valor de $y = g(x)$.

Ejemplo 4 Supongamos que X tiene densidad $p_X(x)$, y sea $Y = X^2$. Aquí, para $y > 0$ hay dos valores de x tales que $y = x^2$, a saber $x = \sqrt{y}$

y $x = -\sqrt{y}$. Si $y < 0$ no hay tales valores de x. Además, la derivada es $dy/dx = 2x$. Entonces

$$p_Y(y) = \frac{p_X(\sqrt{y}) + p_X(-\sqrt{y})}{2\sqrt{y}} \text{ con } y > 0.$$

■

Cambio de variable para $z = g(x, y)$

Supongamos que $Z = g(X, Y)$ en donde $g : \mathbb{R}^2 \to \mathbb{R}$ es una función diferenciable con $\partial_y g \neq 0$.[1] Nuestro objetivo es hallar la densidad de Z en función de la densidad conjunta de X e Y.

[1] Esta condición no es realmente necesaria, pero la asumimos para simplificar la exposición. Basta por ejemplo con asumir que el gradiente de g no se anula en ningún conjunto de área positiva.

Para esto haremos el siguiente truco. Consideremos el cambio de coordenadas en el plano $\mathbb{R}^2$ dado por $(x, y) \mapsto (x, g(x, y)) = (x, z)$. El área de un pequeño rectángulo de vértice (x, y) y lados dx y dy en el plano xy es $dxdy$ (Ver la Figura 6). Un tal rectángulo es enviado por la transformación de coordenadas en un rectángulo de vértice (x, z) y área $dxdz$.

Pero como $dz = |\partial_y g(x, y)| dy$, vemos que la relación de áreas es

$$\frac{dxdy}{dxdz} = \frac{1}{|\partial_y g(x, y)|}.$$

Del mismo modo que hicimos en dimensión 1, se debe cumplir la identidad

$$p(x, z)dxdz = p(x, y)dxdy \quad (\text{con } z = g(x, y))$$

de donde

$$p(x, z) = p(x, y)\frac{dxdy}{dxdz} \quad (\text{con } z = g(x, y)).$$

Usando la relación de áreas que recién describimos, podemos re-escribir esta ecuación como:

$$p(x, z) = \frac{1}{|\partial_y g(x, y)|} p(x, y) \quad (\text{con } z = g(x, y)).$$

Podemos hallar la densidad marginal de Z integrando en X:

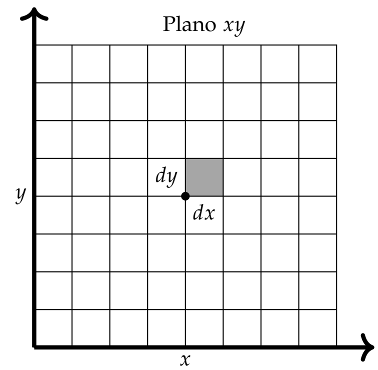
$$p(z) = \int_{-\infty}^{+\infty} p(x, z)dx = \int_{-\infty}^{+\infty} \frac{1}{|\partial_y g(x, y)|} p(x, y)dx.$$

Es decir:

Densidad de $Z = g(X, Y)$

Si $Z = g(X, Y)$, entonces su densidad es

$$p(z) = \int_{-\infty}^{+\infty} \frac{1}{|\partial_y g(x, y)|} p(x, y)dx \quad (\text{con } z = g(x, y)).$$



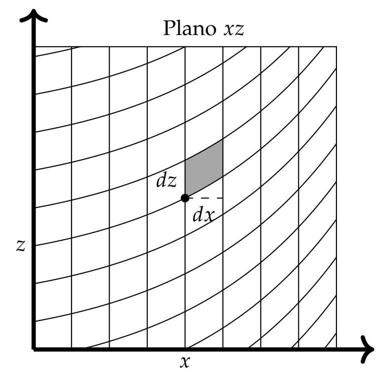


Figura 6: Relación entre el área $dxdz$ y $dxdy$.

Ejemplo 5 — Densidad de la suma de independientes. Supongamos que X e Y son independientes y que $Z = g(X, Y) = X + Y$. Usemos la fórmula anterior para hallar la densidad de Z.

Tenemos que $\partial_y g(x, y) = 1$, de donde

$$p(z) = \int_{-\infty}^{+\infty} p(x, y) dx \text{ (con } z = g(x, y)) = \int_{-\infty}^{+\infty} p(x) p(z - x) dx,$$

en donde en la última igualdad hemos usado que X e Y son independientes, por lo que $p(x, y) = p(x)p(y)$, y que $z = x + y$ por lo que $y = z - x$.

Este caso particular se conoce como fórmula de *convolución*. Es un tema que volveremos a ver más adelante. ■

Ejemplo 6 — Densidad de un producto de independientes. Sean X e Y independientes y $Z = g(X, Y) = XY$. Queremos hallar la densidad de Z.

La derivada parcial $\partial_y g(x, y) = x$, por lo aplicando la fórmula general

$$p(z) = \int_{-\infty}^{+\infty} \frac{1}{|x|} p(x, y) dx \text{ (con } z = xy) = \int_{-\infty}^{+\infty} \frac{1}{|x|} p(x) p(z/x) dx,$$

en donde en la última igualdad hemos usado que X e Y son independientes, por lo que $p(x, y) = p(x)p(y)$, y que $z = xy$ por lo que $y = z/x$.

Por ejemplo, si X e Y tienen distribución uniforme en $(0, 1)$, vemos que su producto Z tiene densidad

$$p(z) = \int_z^1 \frac{1}{x} dx = -\ln(z) \quad (\text{con } 0 < z < 1).$$

En particular

$$E(Z) = -\int_0^1 z \ln(z) dz = \frac{1}{4} = E(X) E(Y).$$

En seguida veremos que esta es una propiedad general de los productos de variables independientes. ■

Esperanza de una función de variables continuas

La aplicación más importante que daremos ahora de la fórmula para la densidad de $Z = g(X, Y)$, es que nos permite probar una ecuación análoga a la (1) para variables continuas.

Esperanza de una función de variables continuas

Sean X e Y dos variables aleatorias continuas, y $g : \mathbb{R}^2 \to \mathbb{R}$ una función. Entonces

$$E\left(g(X,Y)\right) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y)p(x,y)dxdy, \tag{2}$$

en donde $p(x,y)$ es la densidad conjunta de X e Y.

Demostración. Llamemos $Z = g(X,Y)$. Sabemos que la densidad de Z está dada por

$$p(z) = \int_{-\infty}^{+\infty} \frac{1}{|\partial_y g|} p(x,y)dx \text{ (con } z = g(x,y)).$$

Luego, de la definición de esperanza

$$E\left(Z\right) = \int_{-\infty}^{+\infty} zp(z)dz = \int_{-\infty}^{+\infty} z \int_{-\infty}^{+\infty} \frac{1}{|\partial_y g|} p(x,y)dxdz.$$

Remplazando $z = g(x,y)$ obtenemos

$$E\left(Z\right) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{z}{|\partial_y g|} p(x,y)dxdz = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y)p(x,y)\frac{dxdz}{|\partial_y g|}.$$

Recordemos que $|\partial_y g|dy = dz$, por lo que

$$E\left(Z\right) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y)p(x,y)\frac{dxdz}{|\partial_y g|} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y)p(x,y)dxdy.$$

Esto es precisamente lo que queríamos demostrar. □

Propiedades básicas de la esperanza

Varios casos particulares, pero muy importantes, se deducen de las fórmulas (1) y (2).

Esperanza de la suma

Para *todas* X e Y se tiene $E\left(X+Y\right) = E\left(X\right) + E\left(Y\right)$.

Demostración. Supongamos primero que X e Y son discretas. Tomemos $g(x,y) = x + y$ en la fórmula (1). Entonces

$$\begin{aligned} E\left(X+Y\right) &= \sum_{x,y}(x+y)P\left(X=x, Y=y\right) \\ &= \sum_{x,y} xP\left(X=x, Y=y\right) + \sum_{x,y} yP\left(X=x, Y=y\right). \end{aligned}$$

Notar que para cada x fijo, $\sum_{y \in R_Y} P(X = x, Y = y) = P(X = x)$. De aquí resulta que el primer término en la suma anterior es igual a

$$\sum_{x \in R_X} \sum_{y \in R_Y} xP(X = x, Y = y) = \sum_{x \in R_X} xP(X = x) = E(X).$$

Un razonamiento análogo muestra que el segundo término es igual a $E(Y)$.

Veamos ahora cómo es el argumento para el caso en que X e Y son continuas. Usando la ecuación (2) tenemos

$$\begin{aligned}
E(X+Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x+y) p(x,y) dx dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xp(x,y) dx dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yp(x,y) dx dy \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} p(x,y) dy \right) dx + \int_{-\infty}^{\infty} y \left(\int_{-\infty}^{\infty} p(x,y) dx \right) dy \\
&= \int_{-\infty}^{\infty} xp(x) dx + \int_{-\infty}^{\infty} yp(y) dy = E(X) + E(Y).
\end{aligned}$$

Esto termina la demostración. □

Ejemplo 7 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:

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

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

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 (es una propiedad sumamente importante que demostraremos la clase que viene). El valor esperado de X es entonces

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

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

Esperanza del producto

Si X e Y son *independientes* entonces $E(XY) = E(X)E(Y)$.

Demostración. Supongamos primero que X e Y son discretas. Tomemos $g(x,y) = xy$ en la fórmula (1). Entonces

$$\begin{aligned} E(XY) &= \sum_{x,y} xyP(X = x, Y = y) = \sum_{x,y} xyP(X = x)P(Y = y) \\ &= \sum_{x,y} [xP(X = x)] \cdot [yP(Y = y)]. \end{aligned}$$

Pero esta última suma es igual a

$$\left[\sum_x xP(X = x)\right] \cdot \left[\sum_y yP(Y = y)\right] = E(X)E(Y)$$

como queríamos demostrar.

Veamos ahora cómo es el argumento para el caso en que X e Y son continuas. Usando la ecuación (2) tenemos

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (xy)p(x,y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyp(x)p(y)dxdy \\ &= \left(\int_{-\infty}^{\infty} xp(x)dx\right)\left(\int_{-\infty}^{\infty} yp(y)dy\right) = E(X)E(Y). \end{aligned}$$

Esto termina la demostración. □

Ejemplo 8 Volvamos al ejemplo de los cumpleaños para hacer notar una diferencia que tiene el valor esperado respecto de una suma a respecto de un producto.

Consideremos la variable X_{ijk} que vale 1 si la terna $\{i,j,k\}$ cumple el mismo día, y 0 si no. Entonces X_{ijk} es Bernoulli de esperanza $1/365^2$.

El producto $X_{123}X_{124}$ también es Bernoulli, y vale 1 cuando los cuatro 1, 2, 3, y 4 cumplen el mismo día. Entonces

$$\begin{aligned} E(X_{123}X_{124}) &= P(\text{1,2,3,4 cumplen el mismo día}) \\ &= \frac{1}{365^3} \neq E(X_{123})E(X_{124}). \end{aligned}$$

Esto es porque X_{123} y X_{124} no son independientes. ■

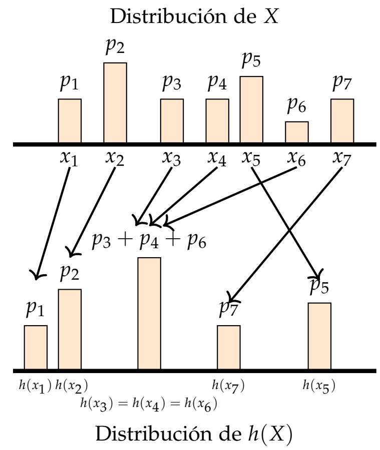


Figura 7: La distribución de $h(X)$ corresponde a cambiar de lugar los bloques, a veces poniendo varios bloques en el mismo lugar. La función h nos indica en dónde colocarlos.

Esperanza de una función de una variable discreta

Si $h : \mathbb{R} \to \mathbb{R}$ y X es discreta, entonces $E\left(h(X)\right) = \sum_x h(x)p(x)$.

Demostración. Basta tomar $g(x, y) = h(x)$ en (1), con $Y = c$ una constante. Entonces

$$E\left(h(X)\right) = E\left(g(X, c)\right) = \sum_x g(x, c)P\left(X = x, Y = c\right) = \sum_x h(x)p(x),$$

como queríamos. Ver la Figura 7 para una ilustración de ésta fórmula con bloques en un tablón. □

Esperanza de una función de una variable continua

Si $h : \mathbb{R} \to \mathbb{R}$ y X continua, entonces $Eh(X) = \int_{-\infty}^{\infty} h(x)p(x)dx$.

Demostración. Sea $Y = h(X)$. Por simplicidad, supongamos que h es creciente. El caso general es análogo pero más engorroso de escribir. Por la fórmula de cambio de variable en una dimensión, sabemos que

$$p(y) = \frac{1}{h'(x)}p(x) \quad (\text{con } y = h(x)).$$

Entonces

$$E\left(Y\right) = \int_{-\infty}^{\infty} yp(y)dy = \int_{-\infty}^{\infty} y\frac{1}{h'(x)}p(x)dy.$$

Usando que $dy = h'(x)dx$ y que $y = h(x)$, obtenemos

$$E\left(Y\right) = \int_{-\infty}^{\infty} h(x)p(x)dx$$

como queríamos demostrar. □

Las constantes salen para afuera

Si c es una constante entonces $E\left(cX\right) = cE\left(X\right)$.

Demostración. Si X es discreta, poniendo $h(x) = cx$ en la fórmula anterior,

$$E\left(cX\right) = \sum_{x \in R_X} cxP\left(X = x\right) = c\sum_{x \in R_X} xP\left(X = x\right) = cE\left(X\right),$$

obtenemos lo que queríamos demostrar. El caso continuo es análogo cambiando suma por integral. □

Ejemplo 9 Sea X el resultado de lanzar un dado, y sea $Y = X^2$. Calcular $E(Y)$.

Como los valores son pocos, podemos hacer una tabla

X	1	2	3	4	5	6
Y	1	4	9	16	25	36
prob	1/6	1/6	1/6	1/6	1/6	1/6

Notar que en este caso la probabilidad para cada valor de Y es la misma que la del correspondiente valor de X. Esto es porque $h(x) = x^2$ es inyectiva en $\{1,2,3,4,5,6\}$.

La esperanza es entonces

$$E(Y) = E\left(X^2\right) = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + \ldots + 6^2 \cdot \frac{1}{6} = 15.167$$

Es el mismo valor que obtendríamos aplicando la fórmula (1). ■

Ejemplo 10 Sea g la función $g(x) = x^2$. Consideremos una variable X con distribución uniforme en los enteros $\{-n,\ldots,n\}$. ¿Cuál es la distribución de $g(X)$?

Los valores posibles que puede tomar $g(X)$ son los cuadrados

$$R_{X^2} = \{0,1,4,\ldots,n^2\}.$$

¿Y con qué probabilidad los toma? La variable toma el valor k^2 cuando X toma uno de los valores $-k$ y k. Como X toma cada uno de sus valores con probabilidad $1/(2n+1)$, vemos que

$$\begin{aligned} p_{X^2}(k^2) &= P\left(X^2 = k^2\right) = P(X=-k) + P(X=k) \\ &= p_X(-k) + p_X(k) = \frac{2}{2n+1}. \end{aligned}$$

La esperanza de X es igual a 0. Sin embargo, la esperanza de X^2 es

$$\begin{aligned} E\left(X^2\right) &= \sum_{k=0}^{n} k^2 p_{X^2}(k^2) = \frac{2}{2n+1}\sum_{k=0}^{n} k^2 \\ &= \frac{2}{2n+1} \cdot \frac{n(n+1)(2n+1)}{6} = \frac{n(n+1)}{3}. \end{aligned}$$

Notar que, en particular, $E\left(X^2\right) \neq E(X)^2$.

Si miramos con detalle la cuenta anterior, vemos que hemos probado que

$$E\left(X^2\right) = \sum_{k=-n}^{n} k^2 p_X(k),$$

que no es otra cosa que la fórmula (1). ■

Ejemplo 11 Se lanzan dos dados y X representa la suma de los resultados. Supongamos que las ganancias de una determinada apuesta están representadas por la variable $Y = X^2 - 6X + 1$. ¿Es una buena apuesta?

Debemos calcular la ganancia esperada $E(Y)$. Usando la fórmula

$$E(Y) = \sum_{j=2}^{12} \left(j^2 - 6j + 1\right) p(j)$$

en donde $p(j) = P(X = j)$ es la f.p.p. de X, que se muestra en la tabla siguiente:

X	2	3	4	5	6	7	8	9	10	11	12
Y	-7	-8	-7	-4	1	8	17	28	41	56	73
prob	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

De aquí resulta que $E(Y) = 83/6 = 13.833$.

Dejamos para que verifiquen que se obtiene el mismo resultado aplicando la definición de valor esperado a la variable Y. ■