Unidad 2. Variables aleatorias discretas

El siguiente material es sólo orientativo, está sujeto a correcciones y se irá completando con el intercambio en el aula.

1. Conceptos generales

Definición 1. *Una variable aleatoria X es una función definida sobre el espacio muestral Ω de un espacio de probabilidad.*

La noción de variable aleatoria cobra sentido probabilístico cuando se introduce la noción de *función de distribución*, dada a continuación.

Definición 2. *Dada una variable aleatoria X llamaremos función de distribución (asociada a X) a la función con dominio en $\mathbb{R}$ y a valores en el intervalo $[0,1]$ definida como*

$$F(x) = P\left(\{\omega \in \Omega : X(\omega) \leq x\}\right) = P(X \leq x), \forall x \in \mathbb{R}.$$

Cuando sea necesario, para indicar la dependencia de F respecto de la variable aleatoria X utilizaremos la notación $F_X(x)$.

Notar que al evento $\{\omega \in \Omega : X(\omega)\}$ se lo denota con $(X \leq x)$. Salvo mención contraria se utilizará esta notación simplificada.

Problema 1. *Para la variable aleatoria X : "número de caras en el lanzamiento de tres monedas (equilibradas)" hallar la función de distribución de X.*

Notar que el espacio muestral es $\Omega = \{(ccc), (ccx), \ldots, (xxx)\}$ y la imagen de la variable aleatoria es el conjunto $\{0, 1, 2, 3\}$. Fijemos algunos valores x y calculemos $F(x)$:

- si $x = -1/2$,

$$\begin{aligned} F(x) &= F(-1/2) \\ &= P(X \leq -1/2) \\ &= P(\text{número de caras sea menor o igual a } -1/2) \\ &= P(\emptyset) = 0. \end{aligned}$$

  Más generalmente, cualquiera sea $x < 0$, $F(x) = 0$.

- si $x = 1/4$,

$$\begin{aligned} F(x) &= F(1/4) \\ &= P(\{(xxx)\}) \\ &= P(X = 0) \\ &= 1/8 \end{aligned}$$

Y para cualquier número x en el intervalo $[0, 1)$, $F(x) = 1/8$

De este modo podemos continuar el análisis y arribar a la expresión general de F:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1/8 & \text{si } 0 \leq x < 1 \\ 4/8 & \text{si } 1 \leq x < 2 \\ 7/8 & \text{si } 2 \leq x < 3 \\ 1 & \text{si } x \geq 3 \end{cases}$$

En la Figura 1 se representa el gráfico de F[1] donde los círculos blancos marcan puntos excluidos.

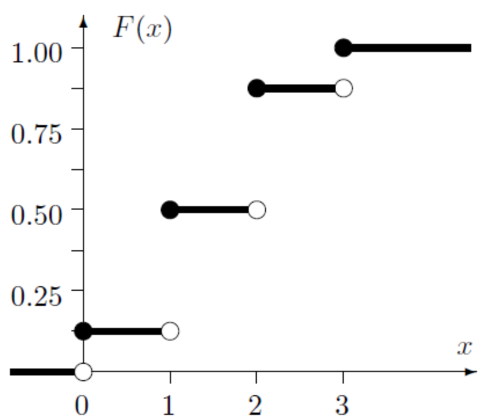

Figura 1: Función de distribución de X : "número de caras en tres lanzamientos de una moneda"

$\square$

Tener en cuenta que si definimos la variable aleatoria Y : "número de cruces en el lanzamiento de tres monedas" entonces $F_Y(z) = F_X(z)$ para todo z en $\mathbb{R}$. Es decir que las variables como funciones pueden ser distintas pero las funciones de distribución pueden ser idénticas. Este hecho es importante ya que, en general, no interesa tanto el dominio de una variable aleatoria X ni su carácter funcionalfunción de distribución .

El siguiente teorema caracteriza a cualquier función de distribución.

Teorema 1. *Para una variable aleatoria X con función de distribución F son válidas las siguientes afirmaciones:*

a) $\forall x, y \in \mathbb{R}$: *si* $x < y$ *entonces* $F(x) \leq F(y)$.

b) $\lim_{x \to +\infty} F(x) = 1$ *y* $\lim_{x \to -\infty} F(x) = 0$.

[1]La gráfica está tomada del texto de Baron (2014)

c) F es una función continua por derecha.

La propiedad a) afirma que F es monótona creciente.

Se puede probar que vale la "recíproca" del Teorema 1 en el siguiente sentido: "si $F : I\!R \to I\!R$ es una función que satisface a), b) y c) entonces existe un espacio de probabilidad $(\Omega, \mathcal{F}, P)$ y una variable aleatoria X tal que $F_X = F$". Es por ello que afirmamos que el teorema anterior provee una caracterización del concepto de función de distribución .

La siguiente proposición, cuya demostración se deja como ejercicio, da otras propiedades de una función de distribución.

Proposición 1. *Para una variable aleatoria X con función de distribución F se cumplen las propiedades siguientes:*

a) $\forall x \in I\!R : P(X > x) = 1 - F(x)$.

b) $\forall x, y \in I\!R :$ *si* $x < y$ *entonces* $P(x < X \leq y) = F(y) - F(x)$.

c) $\forall x \in I\!R : P(X = x) = F(x) - \lim\limits_{y \to x^-} F(y)$.

Demostración. Ver el apéndice.

Ejemplo 1. *Continuación del Problema 1*

La probabilidad de que el número de caras sea mayor a 1, se puede calcular de acuerdo a la proposición anterior como

$$P(X > 1) = 1 - F(1) = 1 - 4/8.$$

Y

$$P(X = 1) = F(1) - \lim_{y \to 1^-} F(y) = 4/8 - 1/8 = 3/8,$$

que a su vez se puede computar como

$$P(\{(cxx), (xcx), (xxc)\}) = 3/\#\Omega = 3/8$$

utilizando las nociones de la Unidad I. □

2. Variables aleatorias discretas

Definición 3. *Una variable aleatoria X se llama discreta si toma valores en un conjunto finito o infinito numerable.*

En términos más rigurosos, una variable aleatoria X se dice *discreta* si existe un conjunto $A \subset I\!R$ finito o infinito numerable tal que

$$P(X \in A) = 1$$

(ver las Notas de Clase de Yohai, 2006, pp. 41–43).

Definición 4. *Dada una variable aleatoria discreta X se llama rango de X a*

$$R_X = \{x \in I\!R : P\,(X = x) > 0\}\,.$$

Se puede demostrar que

$$P\,(X \in R_X) = 1$$

y además es el menor conjunto a lo sumo numerable con esta propiedad; i.e. que si conjunto B es a lo sumo numerable con $P(X \in B) = 1$ entonces $R_X \subset B$. Denotemos entonces $R_X = \{x_1, x_2, \ldots\}$.

Definición 5. *Dada una variable aleatoria discreta X a la función $f : I\!R \to [0, 1]$ definida como*

$$f(x) = P(X = x)$$

se la denomina función de probabilidad de masa

Dada una variable aleatoria discreta denominamos *distribución de la variable* ya sea a su función de probabilidad de masa o a su función de distribución. Notar que conociendo una u otra función queda determinado el rango de la variable.

De aquí en más en esta unidad las variables aleatorias que consideraremos serán discretas.

Varios "hechos" importantes se resumen en la siguiente proposición.

Proposición 2. *Sea X una variable aleatoria discreta con función de probabilidad de masa f y función de distribución F. Entonces son válidas las siguientes tres afirmaciones.*

i) $\forall x \in I\!R$: *$f(x) > 0$ si y sólo si $x = x_i$ para algún x_i en R_X.*

ii)

$$\sum_{x \in R_X} f(x) = \sum_{i=1}^{\infty} f(x_i) = 1.$$

iii) F es discontinua en x si y sólo si $x = x_i$ para algún x_i en R_X.

iv) $\forall x \in I\!R$:

$$F(x) = \sum_{i : x_i \leq x} f(x_i).$$

Demostración: ejercicio.

Observamos que

- El Teorema 1 afirma que una función de distribución F cualquiera es monótona creciente y por iii) de la Proposición 2 si la variable aleatoria es discreta entonces F es una función escalera con discontinuidades a salto finito en los puntos del rango R_X y la magnitud del salto en un punto x_i es

$$\lim_{y \to x_i^+} F(y) - \lim_{y \to x_i^-} F(y) = F(x_i) - \lim_{y \to x_i^-} F(y) = f(x_i).$$

Esta igualdad significa que en cada x_i del rango hay acumulada una "masa" de magnitud $f(x_i)$.

- El ítem i) de la Proposición 2 nos dice que efectivamente el rango de la variable aleatoria es el soporte de la función de probabilidad de masa f.

Ejemplo 2. *Continuación del Problema 1*

La siguiente tabla

x	$f(x)$
0	1/8
1	3/8
2	3/8
3	1/8

nos da la función de probabilidad de masa de X. El rango de esta variable es $R_X = \{0, 1, 2, 3\}$.

□

La noción de variable aleatoria se generaliza del siguiente modo.

Definición 6. *Dadas X e Y variables aleatorias discretas al par (X, Y) se le llama vector aleatorio discreto y a la función $f : \mathbb{R}^2 \to [0, 1]$ dada por*

$$f(x, y) = P(X = x, Y = y)$$

se le llama función de probabilidad de masa conjunta. A

$$F(x, y) = P(X \leq x, Y \leq y), \forall (x, y) \in \mathbb{R}^2$$

se le denomina función de distribución conjunta. A f_X y F_X se le llaman función de probabilidad de masa marginal y función de distribución marginal, respectivamente (y análogamente para f_Y y a F_Y).

Análogamente a como hicimos en dimensión 1 se define el rango $R_{(X,Y)}$. La Definición 6 se generaliza a p dimensiones con $p > 2$.

Supongamos que conocemos el rango $R_{(X,Y)}$ y la función de probabilidad de masa de un vector (X, Y) (o su función de distribución), entonces es posible hallar las distribuciones marginales del siguiente modo:

$$\forall x, y \in I\!R: \ f_X(x) = P(X = x) = \sum_y f(x, y) \quad (1)$$

$$f_Y(y) = P(Y = y) = \sum_x f(x, y). \quad (2)$$

¿Y si conocemos las distribuciones marginales podemos reconstruir la conjunta?.

Ejemplo 3. *Un programa consiste de dos módulos. El número de errores en el primer módulo es una variable aleatoria X y el número de errores en el segundo módulo es otra variable aleatoria Y. Supongamos que la función de probabilidad de masa conjunta es*

$$f(0,0) = f(0,1) = f(1,0) = 0.2, \ f(1,1) = f(1,2) = f(1,3) = 0.1 \ y \ f(0,2) = f(0,3) = 0.05.$$

Observar entonces que la función de probabilidad de masa marginal de X es $f_X(0) = f_X(1) = 1/2$ y la de la Y es $f_Y(0) = 0.4$, $f_Y(1) = 0.3$ y $f_Y(2) = f_Y(3) = 0.15$.

3. Esperanza y Varianza

Dada una variable aleatoria X introducimos a continuación conceptos que tendrán importancia tanto en probabilidad como en estadística.

Definición 7. *Si X es una variable aleatoria discreta entonces la esperanza se define como*

$$\mathrm{E}(X) = \sum_{x \in R_X} x f(x)$$

si la serie converge absolutamente.

La condición de que la serie converja absolutamente, es decir que $\sum_{i=1}^{\infty} |x_i| f(x_i) < \infty$, se impone cuando el rango R_X es numerable. En el caso que sea finito la condición se cumple trivialmente.

La esperanza es *un promedio ponderado o pesado* de los valores del rango de X donde los *pesos* son los valores de la función de probabilidad de masa . A la esperanza se le denomina también media o promedio de la variable aleatoria X.

Ejemplo 4. *Se lanza una moneda equilibrada y X es la variable aleatoria que vale 0 si sale cruz y 1 en caso contrario. Así, la función de probabilidad de masa de X es*

$$f(0) = f(1) = 1/2 \ y \ f(x) = 0 \ en \ caso \ contrario.$$

Notar que $R_X = \{0, 1\}$ *La esperanza de* X *es*

$$\mathrm{E}\,(X) = \sum_{x \in R_X} x f(x) = 0 \cdot f(0) + 1 \cdot f(1) = f(1) = 1/2.$$

¿Qué ocurre si la moneda no está equilibrada?

Ejemplo 5. *En el contexto del ejemplo anterior supongamos que (la probabilidad de que aparezca cruz es)* $f(0) = 3/4$. *Para este caso* $\mathrm{E}\,(X) = 1/4$.

Un **modelo físico** para estos dos ejemplos es abordado en Baron (2014, págs. 47 y ss).

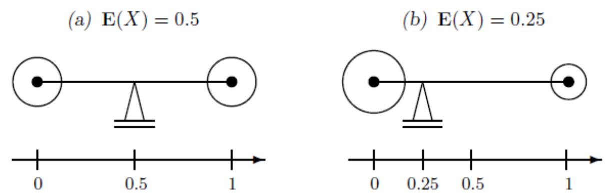

Figura 2: La esperanza como centro de gravedad

En a) de la Figura 2 ponemos dos masas, cada una de 0.5 unidades en los puntos 0 y 1 y los conectamos con con una varilla firme pero sin peso. Las masas representan las probabilidades $f(0)$ y $f(1)$ de la moneda equilibrada. Si buscamos un punto en el cual el sistema esté equilibrado, ese punto está en el medio por la simetría y es el centro de gravedad. En cambio en en b) de la Figura 2 el sistema está balanceado en 0.25 que es su centro de gravedad. En ambas situaciones el centro de gravedad es $\mathrm{E}(X)$.

Problema 2. *Consideremos ahora el experimento de arrojar una moneda sucesivamente hasta que aparezca cara. En cada lanzamiento la probabilidad del evento "cara" es* $p \in (0, 1)$. *Sea* X *la variable que cuenta el número de lanzamientos necesarios hasta que ocurra cara. ¿Cuál es el valor de la esperanza de* X*?*

Solución. Observar que $R_X = \{1, 2, 3, \ldots\} = \mathbb{N}$ y si $x \in R_X$,

$$\begin{aligned} f(x) &= P(X = x) \\ &= P(\text{ que salga cruz en los primeros } x-1 \text{ lanzamientos y cara en el último}) \\ &= (1-p)^{x-1} p, \end{aligned}$$

donde hemos utilizado la independencia.

Calculemos su esperanza. Por definición

$$\mathrm{E}(X) = \sum_{x=1}^{\infty} x p (1-p)^{x-1} = p \sum_{x=1}^{\infty} x r^{x-1} \tag{3}$$

donde en la segunda igualdad $r = 1 - p$.

Y ahora tengamos en cuenta que, intercambiando la sumatoria con la derivada,

$$\sum_{x=1}^{\infty} x r^{x-1} = \sum_{x=1}^{\infty} \frac{d}{dr} r^x = \frac{d}{dr} \sum_{x=1}^{\infty} r^x = \frac{d}{dr} \frac{r}{1-r} = \frac{1}{(1-r)^2}. \tag{4}$$

Y de (3) y (4) de deduce que

$$\mathrm{E}(X) = \frac{1}{p}.$$

Observar que cuando más pequeña sea la probabilidad de cara el número medio de lanzamientos es más grande. □

Analicemos dos propiedades de la esperanza que son simples y conceptualmente importantes. Asumamos que una variable aleatoria X satisface $X \geq 0$, es decir su rango $R_X = \{x_i\}_{i \geq 1}$ es un conjunto de número reales no negativos. Luego $\mathrm{E}(X) = \sum_i x_i f(x_i) \geq 0$ ya que $x_i \geq 0$ por hipótesis y $f(x_i) \geq 0$ por ser f una función de probabilidad de masa .

Supongamos ahora que X vale c con probabilidad 1 siendo c una constante; utilizaremos la notación $X \equiv c$. Luego $R_X = \{c\}$ y entonces $f(c) = P(X = c) = 1$ y entonces claramente $E(X) = c$.

En los dos párrafos anteriores hemos demostrado que la siguiente proposición.

Proposición 3. *Son válidas las siguientes afirmaciones:*

a) Si $X \geq 0$ *entonces* $\mathrm{E}(X) \geq 0$.

b) Si $X \equiv c$ *entonces* $\mathrm{E}(X) = c$.

Consideremos ahora el problema de hallar la esperanza de una función de una variable aleatoria. Supongamos que X es una variable aleatoria con rango $R_X = \{-1, 0, 1\}$ y función de probabilidad de masa constante sobre R_X (i.e. $f_X(x) = 1/3$ si $x \in R_x$). Sea $g(x) = x^2$ y definamos la nueva variable $Y = g(X)$. Nos preguntamos ¿cuál es la esperanza de la variable aleatoria Y?. Si denotamos con f_Y la función de probabilidad de masa de Y entonces $R_Y = \{0, 1\}$ y $f_Y(0) = 1/3$ y $f_Y(1) = 2/3$ con lo cual

$$\mathrm{E}(Y) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}.$$

Observar que podemos re-escribir el valor esperado de Y en términos de la función de probabilidad de masa de X del siguiente modo:

$$\begin{aligned}
\mathrm{E}(Y) = \mathrm{E}[g(X)] &= \sum_{y \in R_Y} y f_Y(y) \\
&= 0 \cdot P(Y = 0) + 1 \cdot P(Y = 1) \\
&= 0 \cdot P(X = 0) + g(1) \cdot P(X = 1) + g(-1) \cdot P(X = -1). \\
&= \sum_{x \in R_X} g(x) f_X(x)
\end{aligned}$$

Esta situación particular se puede generalizar tal como lo enuncia la siguiente proposición.

Proposición 4. ***Ley del inconsciente estadístico***

Dada una variable aleatoria X discreta con función de probabilidad de masa f y sea $g : I\!R \to I\!R$ una función, entonces

$$\mathrm{E}\left[g(X)\right] = \sum_{x \in R_X} g(x) f(x).$$

Demostración. Ver el apéndice.

Esta proposición se puede generalizar al caso multivariado. En efecto, si (X, Y) es un vector aleatorio discreto con función de probabilidad de masa f y $g : I\!R^2 \to I\!R$ entonces

$$\mathrm{E}\left[g(X,Y)\right] = \sum_{(x,y) \in R_{(X,Y)}} g((x,y)) f(x,y). \tag{5}$$

Otras características numéricas importantes de una variable aleatoria se definen a continuación.

Definición 8. *Si X es una variable aleatoria discreta entonces la varianza de X se define como*

$$\mathrm{VAR}(X) = \mathrm{E}\left[(X - \mathrm{E}\,(X))^2\right].$$

y a $\mathrm{DESV}(X) = \sqrt{\mathrm{VAR}(X)}$ se le denomina desvío de la variable aleatoria .

Notar que por la Proposición 4 la varianza se escribe como

$$\mathrm{VAR}\,(X) = \sum_{x \in R_X} (x - \mathrm{E}\,(X))^2 f(x). \tag{6}$$

Continuación de los Ejemplos 4 y 5. Cuando la moneda está equilibrada, utilizando (6), obtenemos

$$\mathrm{VAR}\,(X) = \sum_{x \in R_X} (x - \mathrm{E}\,(X))^2 f(x) = \frac{1}{2} \cdot (0 - \frac{1}{2})^2 + \frac{1}{2} \cdot (1 - \frac{1}{2})^2 = \frac{1}{4}.$$

¿Y cuando no está equilibrada como en el Ejemplo 5? ¿la varianza crece o decrece?.

Y si $f(0) = 1$, ¿a qué es igual $\mathrm{E}(X)$ y $\mathrm{VAR}(X)$? □

La siguiente proposición nos da propiedades elementales e importantes de la esperanza y varianza. Notar que, a diferencia de la varianza, la esperanza es un *operador lineal.*

Proposición 5. *Sean X e Y variables aleatorias discretas, a y b constantes. Entonces son válidas* *las siguientes igualdades:*

$$\begin{aligned} \mathrm{E}(aX+bY) &= a\mathrm{E}(X)+b\mathrm{E}(Y) & (7)\\ \mathrm{VAR}(aX+bY) &= a^2\mathrm{VAR}(X)+b^2\mathrm{VAR}(Y)+2ab\mathrm{E}\left[(X-\mathrm{E}(X))\,(Y-\mathrm{E}(Y))\right] & (8)\end{aligned}$$

Demostración: ejercicio.

Ejemplo 6. *Continuación del Ejemplo 3.*
Si definimos la variable aleatoria $Z = X + Y$, para calcular $\mathrm{E}(Z)$ una opción es hallar f_Z. Otra resolución, más simple, consiste en utilizar la linealidad de la esperanza. En primer lugar notar que

$$\begin{aligned} \mathrm{E}(X) &= 0\cdot 0.5+1\cdot 0.5 = 0.5\\ \mathrm{E}(Y) &= 0\cdot 0.4+1\cdot 0.3+2\cdot 0.15+3\cdot 0.15 = 1.05\end{aligned}$$

y entonces $\mathrm{E}(Z) = \mathrm{E}(X) + \mathrm{E}(Y) = 0.5 + 1.05 = 1.55.$ □

Utilicemos la última proposición para hallar una fórmula muy simple y útil para la varianza:

$$\begin{aligned} \mathrm{VAR}(X) =&\mathrm{E}((X-\mathrm{E}(X))^2) && \text{por definición}\\ =&\mathrm{E}(X^2-2X\mathrm{E}(X)+(\mathrm{E}(X))^2), && \text{desarrollo de cuadrados}\\ =&\mathrm{E}(X^2)+\mathrm{E}(-2X\mathrm{E}(X))+\mathrm{E}((\mathrm{E}(X))^2), && \text{linealidad de la esperanza}\\ =&\mathrm{E}(X^2)-2\left[\mathrm{E}(X)\right]^2+(\mathrm{E}(X))^2, && \text{utilizando (7) y la Proposición 3}\end{aligned}$$

Hemos demostrado la validez del siguiente corolario a la Proposición 5.

Corolario 1. *Dada una variable aleatoria X es válida la siguiente igualdad*

$$\mathrm{VAR}(X) = \mathrm{E}(X^2) - \left[\mathrm{E}(X)\right]^2. \tag{9}$$

Definición 9. *Dadas X e Y variables aleatorias discretas definimos la covarianza entre las variables como*

$$\mathrm{COV}(X,Y) = \mathrm{E}\left[(X-\mathrm{E}(X))\,(Y-\mathrm{E}(Y))\right].$$

Y el coeficiente de correlación entre X e Y se define como

$$\rho = \frac{\mathrm{COV}(X,Y)}{\mathrm{DESV}(X)\mathrm{DESV}(Y)}.$$

Utilizando las propiedades de la esperanza se puede mostrar que es válida la siguiente igualdad

$$\mathrm{COV}(X,Y) = \mathrm{E}\left(XY\right) - \mathrm{E}(X)\mathrm{E}(Y). \tag{10}$$

Definición 10. *Dos variables aleatorias discretas X e Y se dicen no correlacionadas si* $\text{COV}(X, Y) = 0$.

Notar que por (8) de la Proposición 5 se deduce el siguiente:

Corolario 2. *Si dos variables aleatorias discretas X e Y están no correlacionadas entonces*

$$\text{VAR}(X + Y) = \text{VAR}(X) + \text{VAR}(Y). \tag{11}$$

Observación 1. *Es posible mostrar que $-1 \leq \rho \leq 1$. Además $|\rho| = 1$ si y sólo si con probabilidad 1 una variable aleatoria es una función lineal de la otra variable (1 si la pendiente es positiva y -1 si es negativa).*

4. Independencia

Definición 11. *Dadas X e Y variables aleatorias discretas diremos que son independientes si*

$$\forall x, y \in I\!R : P(X = x, Y = y) = P(X = x)P(Y = y)$$

La igualdad dada en la Definición 11 se puede escribir como

$$\forall x, y \in I\!R : f(x, y) = f_X(x)f_Y(y) \tag{12}$$

siendo f la función de probabilidad de masa del vector (X, Y) y f_X y f_Y las marginales.

La definición de independencia se extiende a más de dos variables del siguiente modo.

Definición 12. *Dada una colección arbitraria $\{X_i\}_{i \in I}$ de variables aleatorias discretas se dirán que son independientes si para todo subconjunto $J \subset I$ finito se cumple que*

$$P\left(\forall j \in J : X_j = x_j\right) = \prod_{j \in J} P\left(X_j = x_j\right), \forall x_j \in I\!R, \forall j \in J.$$

La siguiente implicación es de gran importancia.

Proposición 6. *Si X e Y son variables aleatorias independientes entonces*

$$\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y).$$

Demostración: ver apéndice.

La proposición puede enunciarse de un modo más general. Esto es, si g y h son funciones y $g(X)$ y $h(Y)$ son independientes entonces

$$\mathrm{E}(g(X)h(Y)) = \mathrm{E}(g(X))\mathrm{E}(h(Y)).$$

La proposición anterior tiene una consecuencia bien importante.

Corolario 3. *Si X e Y son variables aleatorias independientes entonces están no correlacionadas.*

El siguiente contraejemplo muestra que la recíproca de la proposición anterior es falsa.

Ejemplo 7. *Sea Z una variable aleatoria con $R_Z = \{-1, 1\}$ y $P(Z = 1) = P(Z = -1) = 1/2$. Sea Y una variable aleatoria independiente de Z con $R_Y = \{1, 2\}$ con $P(Y = 1) = P(Y = 2) = 1/2$. Definamos $X = YZ$. Entonces X e Y están no correlacionadas pero no son independientes ya que $P(X = 1, Y = 2) \neq P(X = 1)P(Y = 2)$.*

5. Familias de variables aleatorias discretas

A continuación abordaremos diferentes familias de distribuciones en probabilidad discretas. En algunas se puede “deducir” de los supuestos físicos la ley en probabilidades o sea su distribución.

Distribución binomial de parámetros n y p

Un ensayo de Bernoulli es un experimento con dos resultados posibles, denominados *éxito* y *fracaso*. Consideremos el experimento aleatorio consistente en repetir n ensayos de Bernoulli independientes con probabilidad de éxito p. Sea X la variable definida como el número de éxitos en los n ensayos. Entonces $R_X = \{0, 1, \dots, n\}$ y la función de probabilidad de masa de X es

$$f_X(k) = \binom{n}{k} p^k (1-p)^{n-k},\ k = 0, 1, \dots, n. \tag{13}$$

Para mostrar que (13) es válida notar que el espacio muestral asociado al experimento se puede representar como

$$\Omega = \{(\omega_1, \dots, \omega_n) : \omega_i \in \{0, 1\}\}$$

donde la i-ésima coordenada ω_i vale 1 o 0 si el ensayo arroja éxito o fracaso, respectivamente.

Fijemos ahora $0 \leq k \leq n$ y definamos $A_k = (X = k)$; es decir, el evento que consiste de todas las n-uplas con k unos. Así, si ω^* es un elemento fijo de A_k podemos escribir

$$P(A_k) = P(X = k) = \# A_k P(\{\omega^*\}).$$

Sin pérdida de generalidad elegimos

$$\omega^* = (\underbrace{1, \dots, 1}_{k \text{ unos}}, \underbrace{0, \dots, 0}_{n-k \text{ ceros}}).$$

Utilizando la independencia

$$P(\{\omega^*\}) = \underbrace{p \cdot \ldots \cdot p}_{k \text{ veces}} \cdot \underbrace{1-p \cdot \ldots \cdot 1-p}_{n-k \text{ veces}} = p^k (1-p)^{n-k}. \tag{14}$$

Por lo estudiado en la unidad anterior $\#A_k = \binom{n}{k}$ y de aquí y de (14) se deduce (13).

A cualquier variable X con función de probabilidad de masa dada por (13) le llamamos variable aleatoria binomial de parámetros n y p y abreviaremos $X \sim \text{bin}(n,p)$ o $X \sim \text{binomial}(n,p)$.

A la variable aleatoria $X \sim \text{binomial}(1,p)$ se le denomina de Bernoulli.

Problema 3. *Se sabe que los tornillos producidos por una determinada empresa estarán defectuosos con probabilidad* 0.01 *independientemente uno de otro y, ofrece una garantía de devolución de dinero si como máximo 1 de los 10 tornillos está defectuoso. ¿Qué proporción de paquetes vendidos debe reemplazar la empresa?*

Solución. Si X es el número de tornillos defectuosos en un paquete entonces $X \sim \text{binomial}(10, 0.01)$.

Luego, la probabilidad de que un paquete deba ser reemplazado ($P(X \geq 2)$) es

$$\begin{aligned} 1 - P\{X=0\} - P\{X=1\} &= 1 - \left(\begin{array}{c} 10 \\ 0 \end{array}\right)(0.01)^0(0.99)^{10} - \left(\begin{array}{c} 10 \\ 1 \end{array}\right)(0.01)^1(0.99)^9 \\ &\approx 0.004 \end{aligned}$$

Así, sólo el .4 por ciento de los paquetes deberán ser reemplazados.

Distribución binomial negativa de parámetros m y p

Supongamos que se ejecuta indefinidamente y con independencia un ensayo de Bernoulli hasta que aparezcan m éxitos con $m \geq 1$. Sea X la variable que cuenta el número de repeticiones del ensayo hasta que aparezcan los m éxitos. Con un análisis similar al anterior concluimos que $R_X = \{m, m+1, \ldots\}$

$$f_X(k) = \binom{k-1}{m-1} p^m (1-p)^{k-m}, \; k = m, m+1, \ldots \tag{15}$$

A cualquier variable con este rango y esta función de probabilidad de masa le llamamos variable aleatoria binomial negativa con parámetro p y la denotamos $X \sim \text{BN}(m,p)$.

Problema 4. *En una producción de componentes electrónicos el 5 por ciento son defectuosos. Necesitamos encontrar 12 componentes para nuestras 12 PC. Los componentes son testeados secuencialmente hasta que hallemos 12 no defectuosos. Nos preguntamos ¿cuál es la probabilidad de que tengan que ser testeados más de 15 componentes?*

Solución. Sea X la variable que cuenta el número de componentes chequeados hasta que aparezcan 12 que no sean defectuosos. X tiene una distribución binomial negativa con parámetros $m = 12$ y $p = 0.95$. La pregunta del problema es ¿$P(X > 15) = \sum_{x=16}^{\infty} f(x)$? con f la función de probabilidad de masas de X.

Distribución geométrica de parámetro p

A una variable aleatoria $X \sim \mathrm{BN}(1,p)$ se le llama variable aleatoria geométrica de parámetro p y la denotamos con $X \sim$ geométrica(p).

Distribución de Poisson

A una variable aleatoria X le llamaremos de Poisson de parámetro $\lambda > 0$ si $R_X = \{k \in \mathbb{Z} : k \geq 0\}$ y su función de probabilidad de masa es

$$f_X(k) = \mathrm{e}^{-\lambda}\lambda^k/k!,\ k = 0, 1, \ldots \tag{16}$$

y la denotamos como $X \sim \mathrm{Poisson}(\lambda)$.

Problemas típicos en los que el ajuste de una distribución de Poisson es adecuado son los siguientes:

- Número de errores en una página o grupo de páginas en un libro.
- Número de paquetes de yerba vendidos en un mercado cada día.
- El número de personas ingresando a una oficina postal por día.
- El número de partículas α emitidas en un período fijo de tiempo por un material radioactivo.

Es importante notar que existe la siguiente aproximación de la binomial a la Poisson:

$$\begin{array}{l} \lim_{n\to\infty} \left(\begin{array}{c} n \\ k \end{array} \right) p^k(1-p)^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!} \\ p \to 0 \\ np \to \lambda \end{array}$$

para cualquier k entero no negativo.

Se suele dar la siguiente "regla práctica"

$$\mathrm{binomial}(n,p) \approx \mathrm{Poisson}(\lambda) \text{ cuando } n \geq 30,\ p \leq 0.05,\ np = \lambda. \tag{17}$$

Abordemos a continuación algunas aplicaciones en términos de problemas.

Problema 5. *El 97 por ciento de los mensajes electrónicos son transmitidos sin error y una transmisión es independiente de la otra. De 200 mensajes, ¿cuál es la probabilidad de que al menos 195 sean transmitidos correctamente?*

Solución. La variable X : "número de mensajes transmitidos correctamente de un total de 200" tiene distribución binomial$(200, 0.97)$ y, sea F_X su función de distribución.

Un simple cálculo muestra que $P(X \geq 195) = 0.4432292$. Si quisiéramos utilizar (17) con $\lambda = 200 \cdot 0.97 = 194$ obtendríamos un valor de 0.4809176.

El valor 0.4809176 no es una buena aproximación al calculado con la ditribución exacta, y se debe a que $p = 0.97$ no es pequeño tal como exige (17). Si en cambio definimos Y : "número de mensajes transmitidos incorrectamente de un total de 200", entonces $Y \sim \text{binomial}(200, 0.03)$ y

$$P(X \geq 195) = P(Y \leq 5) = F_Y(5)$$

con F_Y la función de distribución de Y. Utilizando nuevamente la aproximación (17) con Poisson(6) (ya que $\lambda = 6 = 200 \cdot 0.03$) obtenemos una probabilidad de 0.4456796, el cual sí es un valor más cercano al computado con la distribución binomial.

Analicemos ahora un ejemplo más complejo.

Problema 6. *Si en una clase hay n estudiantes ¿cuál es la probabilidad que ninguno de ellos cumplan años el mismo día? y, ¿cuánto debería valer n de tal modo que esa probabilidad esté por debajo de $1/2$?*

Solución. Asumiendo que el año no es bisiesto, hay un total de 365^n cumpleaños posibles en la clase y, por lo tanto, la probabilidad buscada es

$$\frac{(365) \cdot (364) \cdots (365 - n + 1)}{365^n}.$$

En las líneas que siguen argumentaremos, basándonos en la aproximación de Poisson, que si el número n de estudiantes es igual o superior a 23, la probabilidad de que ninguno cumpla años el mismo día permanece por debajo de $1/2$.

Para un par de estudiantes, i, j consideremos el ensayo de Bernoulli consistente en "comprobar si ellos cumplen o no años el mismo día" y, supongamos que el "éxito" es que cumplan el mismo día. Si definimos la variable X número de éxitos en los $\binom{n}{2}$ pares de estudiantes entonces su distribución es "casi"

$$\text{binomial}\left(\binom{n}{2}, \frac{1}{365}\right).$$

El "casi" es porque si bien hay independencia mutua entre ensayos de Bernoulli, los mismos no son independientes.

Si utilizamos la aproximación de la distribución binomial con la distribución Poisson de parámetro

$$\lambda = \binom{n}{2} \frac{1}{365} = \frac{n(n-1)}{730},$$

entonces

$$\begin{aligned} P(\text{ no haya dos personas que cumplan el mismo día}) &= P(X = 0) \\ &= \exp\left(-\frac{n(n-1)}{730}\right) \\ &\leq \frac{1}{2} \end{aligned}$$

es equivalente - tomando logaritmo- a

$$n(n-1) \geq 730 \log(2) \approx 506.$$

Resolviendo para n se obtiene $n \geq 23$.

Distribución hipergeométrica de parámetros D, n, N

Supongamos que una urna tiene N bolas de las cuales hay D negras y $N - D$ blancas. Se extraen secuencialemente de a una n bolas y definimos la variable aleatoria X que cuenta el número de bolas negras extraídas. Observar que el rango de X es

$$R_X = \{k \in \mathbb{N} : \max\{0, n - N + D\} \leq k \leq \min\{n, D\}\}. \tag{18}$$

y que

$$f_X(k) = \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}} \text{ con } k \in R_X \tag{19}$$

Si X es cualquier variable aleatoria con rango (18) y función de probabilidad de masa (19) se le llama variable aleatoria hipergeométrica de parámetros D, n, N, $X \sim$ hipergeométrica(D, n, N) de modo abreviado.

Proposición 7. *Son válidas las siguientes afirmaciones*

a) Si $X \sim$ *binomial*(n, p) *entonces* $\mathrm{E}(X) = np$ *y* $\mathrm{VAR}(X) = np(1-p)$.

b) Si $X \sim$ *Poisson*(λ) *entonces* $\mathrm{E}(X) = \lambda$ *y* $\mathrm{VAR}(X) = \lambda$.

c) Si $X \sim$ *geométrica*(p) *entonces* $\mathrm{E}(X) = 1/p$ *y* $\mathrm{VAR}(X) = (1-p)/p^2$.

d) Si $X \sim$ *BN*(m, p) *entonces* $\mathrm{E}(X) = m/p$ *y* $\mathrm{VAR}(X) = (m(1-p))/p^2$.

e) Si $X \sim$ *hipergeométrica*(D, n, N) *y definamos* $p = D/N$. *Entonces* $\mathrm{E}(X) = np$ *y* $\mathrm{VAR}(X) = np(1-p)\dfrac{(N-n)}{N-1}$.

6. Apéndice

En esta sección hallarán las demostraciones de los teoremas y proposiciones enunciadas arriba.

Demostración del Teorema 1 .

a) Si x e y satisfacen que $x < y$ entonces los eventos $(X \leq x)$ y $(X \leq y)$ cumplen que

$$(X \leq x) \subseteq (X \leq y)$$

y por la monotonía de la probabilidad y la definición de función de distribución se cumple que $F(x) \leq F(y)$.

b) Demostremos $\lim_{x \to +\infty} F(x) = 1$, la otra afirmación se prueba análogamente.

Notar en primer lugar que como F es monótona creciente y está acotada, entonces el $\lim_{x\to\infty} F(x)$ existe. Para demostrar la propiedad es suficiente con probar que para cualquier sucesión monótona creciente $\{x_n\}_{n\geq 1}$ tal que

$$\lim_{n\to\infty} x_n = +\infty$$

se cumple que

$$\lim_{n\to\infty} F(x_n) = 1. \tag{20}$$

Si definimos la sucesión de eventos $A_n = (X \leq x_n)$, $n = 1, 2, \ldots$, observemos que

$$\bigcup_{n\geq 1} A_n = \Omega.$$

En efecto, como Ω es el espacio muestral contiene a cualquier evento, en particular a $\bigcup_{n\geq 1} A_n$. Probemos la inclusión recíproca, para lo cual consideremos $\omega \in \Omega$. Como X es una función, $X(\omega)$ es finito y como $\lim_{n\to\infty} x_n = +\infty$ entonces existe n tal que $X(\omega) \leq x_n$ y así $\omega \in A_n$.

De este modo

$$\begin{aligned} \lim_{n\to\infty} F(x_n) &= \lim_{n\to\infty} P(X \leq x_n) \\ &= \lim_{n\to\infty} P(A_n) \\ &= P(\bigcup_{n\geq 1} A_n) \\ &= P(\Omega) = 1 \end{aligned}$$

donde hemos utilizado la la propiedad de continuidad de la probabilidad .

c) Se deja como ejercicio; como sugerencia analizar la demostración de c) de la Proposición 1.

Demostración de la Proposición 1 .

Las afirmaciones de los incicios a) y b) se dejan como ejercicio. Probemos c).

Sea x un número real cualquiera y considerar la sucesión $h_n = 1/n$ con $n \in \mathbb{N}$. Como F es monótona los límites laterales existen en cualquier punto, en particular $\lim_{y\to x^-} F(y)$ existe.

Por otro lado teniendo en cuenta que

$$\bigcup_{n\geq 1} (-\infty, x - \frac{1}{n}] = (-\infty, x),$$

podemos establecer las siguientes relaciones entre los eventos

$$\{X < x\} = \{X \leq x\} - \{X = x\} = \bigcup_{n \geq 1} \{X \leq x - \frac{1}{n}\}$$

y, aplicando nuevamente la propiedad de continuidad de la probabilidad concluimos que

$$P(X = x) = F(x) - \lim_{y \to x^-} F(y)$$

□

Demostración de la Proposición 4 .

Asumamos, sin pérdida de generalidad, que el rango de X, $R_X = \{x_1, \ldots, x_n\}$ es finito y sea $R_Y = \{y_1, \ldots, y_m\}$ el rango de $Y = g(X)$ el que es igual a la imagen por g de R_X. Sea $f_Y(y) = P(Y = y)$ la función de probabilidad de masa de Y. Notar que $0 < m \leq n$.

Para cada y_j renombremos los x_i's de la premimagen $g^{-1}(\{y_j\})$ como $x_i^{(j)}$, $i = 1, \ldots, n_j$, $j = 1, \ldots, m$.

Observar que los conjuntos

$$\{x_1^{(j)}, \ldots, x_{n_j}^{(j)}\}, j = 1, \ldots, m$$

son disjuntos dos a dos, su unión es igual al rango $\{x_1, \ldots, x_n\}$ y $\sum_{j=1}^{m} n_j = n$.

Así

$$\begin{aligned}
\mathrm{E}(Y) &= \sum_{j=1}^{m} y_j f_Y(y_j) \\
&= \sum_{j=1}^{m} y_j \sum_{i=1}^{n_j} f\left(x_i^{(j)}\right) \\
&= \sum_{j=1}^{m} \sum_{i=1}^{n_j} y_j f\left(x_i^{(j)}\right) \\
&= \sum_{j=1}^{m} \sum_{i=1}^{n_j} g\left(x_i^{(j)}\right) f\left(x_i^{(j)}\right) \\
&= \sum_{i=1}^{n_1} g\left(x_i^{(1)}\right) f\left(x_i^{(1)}\right) + \ldots + \sum_{i=1}^{n_m} g\left(x_i^{(m)}\right) f\left(x_i^{(m)}\right) \\
&= \sum_{i=1}^{n} g(x_i) f(x_i)
\end{aligned}$$

□

Demostración de la Proposición 6 . Sea $f(x, y)$ la función de probabilidad de masa del vector (X, Y) y definamos la función $g(x, y) = xy$. Por la definición de independencia expresada por (12) y por la ley del inconsciente estadístico para dos variables dada en (5) se tiene que

$$\begin{aligned}
\mathrm{E}(XY) &= \mathrm{E}(g(x,y)) \\
&= \sum_x \sum_y g(x,y) f(x,y) \\
&= \sum_x \sum_y xy f_X(x) f_Y(y) \\
&= \sum_x x f_X(x) \left[\sum_y y f_Y(y) \right] \\
&= \mathrm{E}(X)\mathrm{E}(Y)
\end{aligned}$$

como queríamos probar. $\square$