Distribuciones de Probabilidad

En una clase de 30 personas hay 5 hijos únicos, 10 personas que tienen un solo hermano/a, 8 personas que tienen 2 herman@s, 6 personas con 3 y 1 persona que tiene... ¡6! Supongamos que elegimos al azar un alumno/a de esta clase. Si tuvieras que apostar, ¿cuántos hermanos dirías que tiene?

No sé qué dirás tú, pero apostar por un solo hermano es lo más apropiado. Mis probabilidades de acertar serían de 10 casos favorables entre 30 casos posibles, 1/3. ¿Cuál es la probabilidad en cada uno de los casos? Hagamos una tabla:

Nº hermanos	0	1	2	3	4	5	6	**Total**
Personas	5	10	8	6	0	0	1	30
Probabilidad	5/30	10/30	8/30	6/30	0/30	0/30	1/30	30/30

Pasando las fracciones a decimales tenemos:

Nº hermanos	0	1	2	3	4	5	6	**Total**
Probabilidad	0′167	0′333	0′267	0′2	0	0	0′0333	1

Esta tabla representa una **distribución de probabilidad** para la variable aleatoria «número de hermanos». Como es lógico, la suma de las probabilidades debe de ser 1 si la distribución está bien hecha. Imagina que tienes 1 tarta: una distribución de probabilidad es como un «reparto» de la tarta entre distintos valores posibles.

La tabla anterior muestra también se puede ver como **función de probabilidad** $P(x)$, donde x es el número de hermanos. Nos pueden pedir $P(0) = 0'167$, ó $P(5) = 0$. O incluso $P(9) = 0$. Eso sí: esta función sólo admite números naturales, ya que no tiene sentido tener un «número decimal de hermanos». Como hay un salto entre dos valores posibles consecutivos, diremos que la distribución de probabilidad es **discreta**.

La función de probabilidad nos puede dar una gráfica (ver figura 1).

Q1. ¿Cuánto tienen que sumar las probabilidades de *cualquier* distribución de probabilidad?

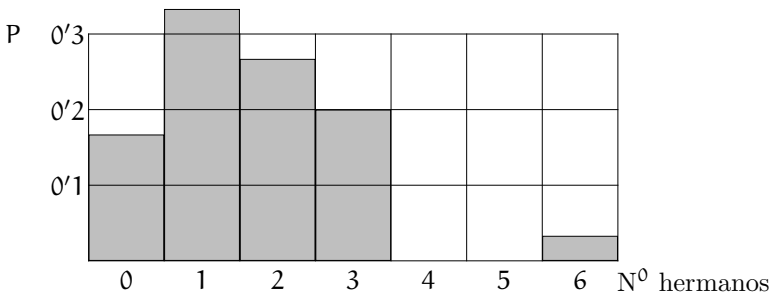

FIGURA 1. Función de probabilidad.

E1. En otra clase hay 10 alumnos de 15 años, 6 de 16 y 4 de 17. Calcula la distribución de probabilidad para la edad en esta clase. Haz la gráfica correspondiente.

Si nos piden la probabilidad de algún suceso más complicado, basta por lo general con sumar valores de la tabla. Por ejemplo, la probabilidad de obtener al azar una persona que tenga 2 ó 3 hermanos es $P(2) + P(3) \approx 0'467$.

E2. En el ejercicio anterior, obtén la probabilidad de cada uno de los siguientes sucesos: $A = \{\text{Tener menos de 17 años}\}$, $B = \{\text{No tener 16 años}\}$.

Llamaremos **función de distribución de probabilidad** a una $F(x)$ que da la probabilidad de obtener un valor que sea *menor o igual* que x. Calculemos la función para el ejemplo de los hermanos. La probabilidad de tener 0 hermanos o menos es, claro, 0′167. La probabilidad de tener 1 hermano (o menos) es la suma de «tener 0» más «tener 1»: $0'167 + 0'333 = 0'5$. Y así sucesivamente obtenemos:

Nº hermanos	0	1	2	3	4	5	6
$F(x)$	0′167	0′5	0′767	0′967	0′967	0′967	1

Te habrás dado cuenta de que cada valor de la tabla es la suma de los que caen por debajo de él en la tabla de la función de probabilidad. El último valor en la tabla de la función de distribución tiene que ser siempre 1, porque la probabilidad de tener un número de hermanos menor o igual que el máximo posible... debe ser 1. También te habrás dado cuenta de que los valores que antes daban 0 ahora se corresponden con valores que «se mantienen», porque no se suma nada. La figura 2 muestra la gráfica de la función de distribución.

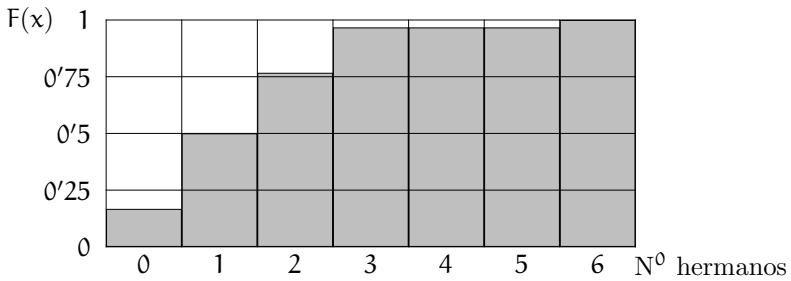

FIGURA 2. Función de distribución de probabilidad.

La propiedad fundamental de la función de distribución que la hace importante es que permite calcular con facilidad la probabilidad de que el resultado caiga en un cierto *intervalo*. Por ejemplo, si nos piden calcular la probabilidad de tener entre 1 y 3 hermanos, tendríamos que sumar:

$$P([1,3]) = P(1) + P(2) + P(3) = 0'8$$

Si fueran muchos los valores a sumar, el cálculo sería muy tedioso. Pero, usando la función de distribución, es más fácil: ¡las sumas ya están hechas! $F(3)$ vale $P(0) + P(1) + P(2) + P(3)$, ¿no? Y $F(0) = P(0)$. Por tanto,

$$p[1,3] = F(3) - F(0) = 0'967 - 0'167 = 0'8$$

Con la función de distribución una resta basta para calcular la probabilidad de cualquier intervalo.

E3. En una clase de bachillerato hay 30 personas. A 6 no les ha quedado ninguna; a 8 les ha quedado una; a 5 les han quedado dos y a 7 les han quedado tres. También hay una persona con 4, una con 5, una con 6 y una con 7. Calcula la función de probabilidad y la función de distribución de probabilidad. Calcula $p[2,5]$ usando primero la función de probabilidad y después la función de distribución.

Q2. Supón que la función $P(x_0) = 0$. ¿Qué ocurrirá con la gráfica de $F(x)$ en x_0? Más difícil: ¿qué ocurriría con la gráfica de $F(x)$ en $x = x_1$ si es allí donde $P(x)$ toma su valor máximo?

Usaremos esta fórmula para expresar la función de distribución en términos de la función de probabilidad:

$$F(x) = \sum_{i \le x} P(i)$$

queriendo decir que para obtener $F(x)$ hay que sumar todos los valores de $P(i)$ con $i \le x$.

E4. Calcula la función de probabilidad y la función de distribución para el suceso aleatorio «tirar un dado». Dibújalas.

E5. Algo más difícil. Calcula la función de probabilidad y la de distribución para el suceso «tirar dos dados y sumar los resultados». Dibújalas también.

Medidas de Centralización.

Imaginad una moneda que tiene por un lado un 7 dibujado y por el otro un 9. Lanzamos la moneda muchas veces y tomamos la media de los resultados que van saliendo. ¿A qué valor nos iremos acercando? Parece intuitivo que será 8 ese valor. Podemos verlo así: el 7 aparece «en promedio» la mitad de las veces, así que contribuye por sí mismo $7 \cdot 1/2 = 3'5$. El 9 aparece la otra mitad de las veces, así que contribuye $9 \cdot 1/2 = 4'5$. Ahora sumamos las dos contribuciones: $3'5 + 4'5 = 8$.

Llamaremos **media** de una variable aleatoria al valor al que tiende el promedio de los resultados de un experimento. Se calcula multiplicando cada valor posible por su probabilidad y sumando los resultados:

$$\mu = \sum_i x_i \cdot P(x_i)$$

Usaremos para ella la letra griega μ. Un ejemplo. Al tirar un dado, ¿a qué valor tiende la media de los resultados? Los valores posibles son del 1 al 6, y todos tienen probabilidad $1/6$, así que:

$$\mu = \langle x \rangle = \sum_i x_i \cdot P(x_i) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3'5$$

E6. Comprueba que el valor esperado para el número de hermanos en el ejemplo del principio de la sección es $\approx 1'667$.

La **moda** de una distribución de probabilidad es simplemente el valor que tiene más probabilidad. En el ejemplo del principio de la sección, es $x = 1$. Por supuesto, puede que el honor de tener el máximo se lo disputen dos valores, y en ese caso se dice que la distribución es *bimodal*. Si son más de dos, se llamará *multimodal*.

La **mediana** de una distribución es el valor x_m que cumple que la probabilidad de obtener un resultado mayor $P(x > x_m)$ es la misma que la de obtener un resultado menor o igual $P(x \leq x_m)$. En el ejemplo de la sección anterior, es $x = 1$, ya que la probabilidad de obtener 0 ó 1 es exactamente $1/2$. Habría sido posible posible que no hubiera ningún valor concreto que cumpla esta condición *exactamente*. En ese caso, habríamos tomado el que más se aproximara.

Notad que, si tenemos la función de distribución, buscar la mediana es muy fácil. No hay más que encontrar el valor x para el que $F(x) = 1/2$. La razón es que la probabilidad de encontrar un valor menor o igual que él será, simplemente, $1/2$. Por tanto, la de encontrar un valor mayor que él también será $1/2$.

Las tres medidas anteriores: media, moda y mediana, se llaman *medidas de centralización* porque se fijan en el «centro» de la distribución. Ojo: no tienen por qué coincidir.

E7. Calcula media (es decir: valor esperado), mediana y moda en la distribución que cumple $P(0) = 0'24$, $P(1) = 0'41$, $P(2) = 0'26$, $P(3) = 0'08$ y $P(4) = 0'01$.

Q3. Inventa una distribución en la que la moda sea mayor que la media y otra en la que ocurra lo contrario. ¿Qué observas?

Medidas de Dispersión I.

«La estadística es aquella ciencia que, si tú tienes una gallina y yo no tengo ninguna, finge repartirlas diciendo que los dos tenemos media gallina.» Esta frase no es cierta del todo. Las medidas de centralización se fijaban en los promedios, pero hay otro tipo de medidas, llamadas de *dispersión*, que estudian cómo de repartida está la probabilidad.

Un ejemplo. Considera la distribución de probabilidad: $P(0) = 0'01$, $P(1) = 0'01$, $P(2) = 0'96$, $P(3) = 0'01$ y $P(4) = 0'01$. Está claro que la probabilidad está muy concentrada en el valor central $x = 3$. Se dice que tiene baja dispersión. En cambio, la distribución $P(0) = 0'25$, $P(1) = 0'25$, $P(2) = 0'25$ y $P(3) = 0'25$ tiene alta dispersión.

Una primera medida es la distancia entre el mayor valor posible y el menor. Esta medida se conoce como **rango** de una distribución. Así, p.ej., en el ejemplo de la primera sección, el número de hermanos iba desde 0 hasta 6, así que el rango es $6 - 0 = 6$. Pero esta medida es muy burda: no tiene en cuenta el hecho de que el 6 es mucho menos probable que el 0...

El valor esperado al lanzar un dado es $3'5$, vale. Pero a veces saldrá un 1 y a veces un 6. ¿Cuál es el valor esperado de la *desviación* respecto de la media?

Veamos: el 1 llevaría una desviación de $(1 - 3'5) = -2'5$ respecto de la media. El 2 llevará una desviación de $(2 - 3'5) = -1'5$ y así sucesivamente hasta el 6, que lleva una desviación de $(6 - 3'5) = 2'5$. La media de estas desviaciones es:

$$\sum (x_i - \mu) \cdot P(x_i) =$$
$$= (-2'5) \cdot \frac{1}{6} + (-1'5) \cdot \frac{1}{6} + (-0'5) \cdot \frac{1}{6} + (0'5) \cdot \frac{1}{6} + (1'5) \cdot \frac{1}{6} + (2'5) \cdot \frac{1}{6} = 0$$

¡Da cero! Claro: hay tantos valores que desvían por arriba como los que desvían por abajo, y al final se cancelan los resultados. ¿Qué podemos hacer para tener una medida de la desviación?

Valor Esperado.

En el último párrafo de la sección anterior hemos calculado la «media de las desviaciones» y comprobado que da cero. Haremos mucho operaciones similares, así que nos compensa inventar una notación especial. Cuando metamos una expresión cualquiera X entre corchetes angulares querremos decir lo siguiente:

$$\langle X \rangle = \sum_i X_i \cdot P(x_i)$$

Llamaremos a $\langle X \rangle$ el **valor esperado** de X. Así, por ejemplo, $\langle x \rangle$ es igual a la media μ. En otras palabras: el valor esperado de la variable es la media.

¿Qué significado tiene $\langle x - \mu \rangle$? Es el valor esperado de las desviaciones respecto de la media, que hemos comprobado que es cero.

Q4. Para una distribución cualquiera, ¿qué significa y cuánto vale $\langle 1 \rangle$?

Medidas de Dispersión II.

Volvamos al problema original de encontrar una medida apropiada de la dispersión de una distribución de probabilidad.

Hay muchas posibilidades. La más sencilla es poner un «valor absoluto» a cada uno de los sumandos para que no se anulen entre sí. En ese caso obtenemos una medida de la desviación:

$$\delta_m = \langle |x - \mu| \rangle = \sum_i |x_i - \mu| \cdot P(x_i) =$$

$$= 2'5 \cdot \frac{1}{6} + 1'5 \cdot \frac{1}{6} + 0'5 \cdot \frac{1}{6} + 0'5 \cdot \frac{1}{6} + 1'5 \cdot \frac{1}{6} + 2'5 \cdot \frac{1}{6} = 1'5$$

Pero no será ésta la medida más *útil* de la desviación. Por razones que veremos en su momento, es más importante la **varianza**, que es el valor esperado de los *cuadrados* de las desviaciones respecto de la media. Como está al cuadrado, se escribe σ^2:

$$\sigma^2 = \left\langle (x - \mu)^2 \right\rangle = \sum_i (x_i - \mu)^2 \cdot P(x_i) =$$

$$= (-2'5)^2 \cdot \frac{1}{6} + (-1'5)^2 \cdot \frac{1}{6} + (-0'5)^2 \cdot \frac{1}{6} + (0'5)^2 \cdot \frac{1}{6} + (1'5)^2 \cdot \frac{1}{6} + (2'5)^2 \cdot \frac{1}{6} \approx 2'9167$$

Pero esta varianza es, hasta cierto punto, «trampa»: por quitar el signo hemos elevado al cuadrado. Todo estará mejor si tomamos al final la raíz cuadrada, ¿no? A ese número le llamaremos **desviación típica** o *desviación cuadrática media*: $\sigma = \sqrt{\sigma^2}$. En nuestro ejemplo, $\sigma \approx \sqrt{2'9167} \approx 1'708$.

Hay una fórmula más útil para la varianza, pero para deducirla necesitamos conocer tres propiedades de los valores esperados (de los corchetes angulares, vamos):

$$\langle kX \rangle = k \langle X \rangle \qquad \langle X_1 \pm X_2 \rangle = \langle X_1 \rangle \pm \langle X_2 \rangle \qquad \langle k \rangle = k$$

donde k es cualquier número (¡no una variable aleatoria!) y X, X_1 y X_2 son cualquier par de expresiones que tengan en sus tripas a la variable aleatoria x.

Q5. Demuestra las tres expresiones antedichas. Por cierto: las dos primeras se suelen resumir diciendo que el valor esperado es *lineal*.

Usando las fórmulas anteriores deducimos:

$$\sigma^2 = \left\langle (x - \mu)^2 \right\rangle = \left\langle x^2 - 2x\mu + \mu^2 \right\rangle = \left\langle x^2 \right\rangle - \langle 2x\mu \rangle + \left\langle \mu^2 \right\rangle =$$
$$\left\langle x^2 \right\rangle - 2\mu \langle x \rangle + \mu^2 = \left\langle x^2 \right\rangle - 2\mu^2 + \mu^2 = \left\langle x^2 \right\rangle - \mu^2.$$

De manera que $\sigma^2 = \left\langle x^2 \right\rangle - \mu^2$, fórmula que es más fácil de emplear.

E8. Calcula el rango, desviación media, varianza y desviación típica para la distribución del ejercicio **E7**.

E9. Supón que generamos cuadrados de manera aleatoria. El lado del cuadrado l se puede considerar una variable aleatoria que sigue la distribución $P(l)$ con $P(2\,\text{cm}) = 0'4$, $P(3\,\text{cm}) = 0'5$, $P(4\,\text{cm}) = 0'1$. Calcula:

a) Valor esperado del lado del cuadrado $\langle l \rangle$. Estima el área esperada elevando este número al cuadrado.

b) Valor esperado del área $\left\langle l^2 \right\rangle$. ¿Coincide con el resultado de la pregunta anterior? ¿Cuál de los dos valores es «mejor»? ¿Por qué?

E10. Considera la distribución de probabilidad correspondiente al lanzamiento de un dado. Calcula el valor esperado de la raíz cuadrada del resultado: $\left\langle \sqrt{x} \right\rangle$.

Distribución de una Suma.

Supongamos que tenemos una variable aleatoria x que sigue una distribución de probabilidad P_x, y otra variable aleatoria y que sigue una distribución de probabilidad P_y. Conocemos sus medias (μ_x y μ_y) y sus varianzas (σ_x^2 y σ_y^2) y nos dicen que ambas variables son totalmente independientes. Nos preguntamos por la media y varianza de una nueva variable: $z = x + y$.

Parece bastante obvio que la media para z será la suma de las medias: $\mu_z = \mu_x + \mu_y$. ¿Qué ocurrirá con las varianzas? Aquí viene la magia: la varianza también es *aditiva*: $\sigma_z^2 = \sigma_x^2 + \sigma_y^2$. Ésta es la razón de que nos importe más la varianza que cualquier otra medida posible de la desviación (como, p.ej., la que usa valores absolutos).

E11. Calcula la media y desviación típica para el experimento: «lanzar dos dados y sumar sus valores». Comprueba que se cumple la ley de suma para medias y varianzas.

Date cuenta de que es la varianza y no la desviación la que se puede sumar. Por tanto, si sumamos N variables aleatorias independientes con la misma distribución:

$$X = \sum_{i=1}^{N} x_i \qquad \Longrightarrow \qquad \mu_X = N\mu_x \qquad \sigma_X^2 = N\sigma_x^2$$

E12. Tiramos un dado veinte veces y sumamos los resultados. Calcula la media y desviación típica del resultado.

La Distribución Binomial.

En el resto de este capítulo vamos a estudiar una distribución de probabilidad particular, pero muy importante. Considera el experimento consistente en «lanzar una moneda N veces y contar el número de caras». Para cada valor de N tenemos una distribución de probabilidad. Por ejemplo, si $N = 1$, entonces la distribución es simplemente $P(0) = 1/2$ y $P(1) = 1/2$.

Q6. Escribe la distribución completa de probabilidad para $N = 2$, $N = 3$ y $N = 4$. Haz las gráficas asociadas.

Llamemos «+» a obtener cara y «−» a obtener cruz, como normalmente. Supongamos que nos piden la probabilidad de lanzar una moneda 5 veces y obtener, exactamente, $+--+-$. Si la moneda no está cargada, la probabilidad de cara y cruz son $P(+) = P(-) = 1/2$. Supondremos siempre que cada lanzamiento es independiente, de forma que:

$$P(+--+-) = \left(\frac{1}{2}\right)^5 = \frac{1}{32}$$

Pero supongamos que no nos interesa esa configuración particular $+--+-$, sino que queremos saber la probabilidad de obtener 2 caras. Aquí las dos caras son en el primer lanzamiento y el el cuarto, pero podrían estar en cualquier otro lugar... ¿cuántas maneras hay de elegir los lugares para las dos caras? Intentemos escribirlos todos:

$$\begin{array}{lllll} ++--- & +-+-- & +--+- & +---+ & -++-- \\ -+-+- & -+--+ & --++- & --+-+ & ---++ \end{array}$$

Son 10 formas. ¿Podríamos haberlo deducido sin escribirlas? Sí. Coincide con el número de maneras de elegir dos miembros de entre un grupo de cinco sin importar el orden. Es decir, el número de combinaciones de 5 elementos tomados de 2 en 2, $C_{5,2}$:

$$C_{5,2} = \binom{5}{2} = \frac{5!}{2!\,3!} = 10$$

En general, la probabilidad de obtener x caras al lanzar una moneda N veces es:

$$P(x) = \binom{N}{x}\left(\frac{1}{2}\right)^5$$

Ésta es un tipo particular de distribución binomial, que veremos en general en el siguiente párrafo.

E13. Dibuja las distintas configuraciones posibles para 6 monedas. Comprueba que, en número, son $C_{6,2}$.

E14. Dibuja la gráfica de la distribución de probabilidad del número de caras para $N = 6$.

E15. Calcula el valor esperado para el número de caras lanzando una moneda 6 veces.

¿Qué ocurre cuando la moneda está cargada? Entonces a la probabilidad de sacar cara la llamaremos p, y llamaremos q a la probabilidad de sacar cruz. Como es lógico, $p + q = 1$. ¿Cuál es la probabilidad de la secuencia $+--+-$? Pues es:

$$P(+--+-) = p \cdot q \cdot q \cdot p \cdot q = p^2q^3$$

es decir, es la misma para cualquier configuración siempre que el número de caras y cruces sea el mismo. Lanzando una moneda N veces, cualquier configuración con x caras tendrá una probabilidad de:

$$p^x q^{N-x}$$

Pero el número de configuraciones con x caras de un total de N lanzamientos es el mismo que antes: $C_{N,x}$. Por tanto, la distribución de probabilidad para el número de caras de una moneda cargada es:

$$P(x) = \binom{N}{x} p^x q^{N-x}$$

Ésta es la **distribución binomial**[1]. Comprueba que, si la moneda es «justa», obtenemos el resultado anterior.

E16. Calcula la distribución de probabilidad para el número de caras al lanzar 4 monedas si la probabilidad de obtener cara es de $0'2$.

Pero, esto de la binomial, ¿sólo sirve para monedas? No. En general se utiliza en *experimentos dicotómicos repetidos* que sean independientes (toma palabreja). Un experimento es «dicotómico» cuando sólo hay dos resultados posibles: cara–cruz, éxito–fracaso, funciona–no funciona, chico–chica, marciano–terrícola... Lo de «repetidos» es fácil de entender, y lo de «independientes» es porque es necesario que el resultado de un experimento no influya en el siguiente (es decir: la moneda no tiene memoria). Veamos unos cuantos ejemplos de su uso:

E17. Los señores de Martínez son premio nacional de natalidad con sus 20 hijos. Calcula la probabilidad de que sean, exactamente, 10 chicos y 10 chicas.

E18. Un bailarín borracho da, a cada palmada de sus borrachos amigos, un paso hacia delante o hacia atrás con la misma probabilidad. Suponiendo que cuando da un paso no se acuerda de cuál fue el anterior (hipótesis bastante aceptable dado su nivel de alcohol), se pide la distribución de probabilidad para su distancia al punto de partida tras 4 pasos.

E19. En una fábrica de coches la probabilidad de que uno salga defectuoso es de $0'1$. Calcula la probabilidad de que no haya ninguno defectuoso, de que haya 1 o de que haya 2 si se han fabricado 15 coches.

E20. Un dado se ha tirado 7 veces y hemos anotado el número de múltiplos de 3 que se obtienen. Calcula la distribución de probabilidad para este valor.

E21. La probabilidad de que apruebe un examen de matemáticas al azar es de $0'01$. ¿Cuántos exámenes tengo que hacer para que la probabilidad de aprobar *al menos uno* sea mayor que $1/2$?

[1] El nombre de *binomial* es histórico: se debe a la relación entre los números combinatorios y el *binomio de Newton*.

E22. En un instituto hay 1.000 alumnos. Calcula la probabilidad de que haya más de 4 que cumplan años el 29 de agosto.

Media y Varianza de la Distribución Binomial.

Consideremos una distribución binomial con $N = 1$ y probabilidades $p + q = 1$. Tenemos, por tanto, $P(0) = q$ y $P(1) = p$. Vamos a calcular la media:

$$\mu = \langle x \rangle = 0 \cdot q + 1 \cdot p = p$$

También podemos calcular su varianza:

$$\sigma^2 = \left\langle (x - \mu)^2 \right\rangle = (0 - p) \cdot 0 + (1 - p) \cdot p = pq$$

Así que $\mu = p$ y $\sigma^2 = pq$. Cuando tenemos N arbitrario es «como si» sumáramos N variables aleatorias del mismo tipo. Sabiendo que la media y la varianza son aditivas, estas fórmulas se nos generalizan a:

$$\mu = Np \qquad \sigma^2 = Npq$$

E23. Calcula el número esperado de chicas que habrá en una familia que tiene 10 hijos sabiendo que la probabilidad de nacimiento de un varón es de $0'54$. Calcula también la desviación típica.

En muchos casos parecidos al problema anterior se dice que el «número de chicas es 5 ± 1», queriendo decir que la media es 5 y la desviación típica 1. Sabiendo esto, calcula:

E24. Sabiendo que una máquina para fabricar bombillas produce cada día 10 ± 1 bombillas defectuosas, ¿podrías estimar cuántas bombillas produce al día?