Capítulo 2. Variable aleatoria

Índice

1. Variable aleatoria y función de distribución

> *Variable aleatoria.* Una variable aleatoria, X, es una función con dominio Ω y recorrido la recta real.

De acuerdo con esta definición, una variable aleatoria asigna un número real a cada suceso elemental del espacio muestral Ω. Esto permite estudiar dichas variables y establecer relaciones entre ellas utilizando las técnicas del análisis matemático.

Para denominar a las variables aleatorias se utilizan letras latinas mayúsculas mientras que sus posibles valores se indican por letras minúsculas. Por ejemplo X es una variable aleatoria y x uno de sus posibles valores.

Ejemplo 1. Cuatro personas juegan al parchís. El color de la persona ganadora puede ser

$$\Omega = \{\text{Rojo, Verde, Azul, Amarillo}\}.$$

Sobre este espacio muestral puede definirse la variable aleatoria $X = 1, \ldots, 4$, que indica el color ganador. Los valores de esta variable son meras etiquetas que indican los colores, pero no tiene sentido realizar sobre ellos cálculos matemáticos como medias, etc. Se trata de una variable nominal.

EJemplo 2. Un conductor tiene que pasar por un semáforo y no sabe si lo encontrará en rojo. El experimento aleatorio tiene el siguiente espacio muestral:

$$\Omega = \{\text{Rojo, Amarillo, Verde}\}$$

La variable $X \equiv$ color del semáforo puede definirse del siguiente modo: $X = 0$ si el semáforo está en rojo, $X = 1$ si está en amarillo y \$X = \$ 2 si está en verde. Se trata de una variable discreta que solo toma tres valores.

En los ejemplos 1 y 2 no existe una correspondencia natural entre los sucesos elementales y los valores numéricos asignados, estos últimos simplemente se utilizan para representar los sucesos y distinguirlos unos sucesos de otros. En los siguientes ejemplos si que existe una forma natural de asignar números a los sucesos elementales.

Ejemplo 3. Supongamos que se define el experimento aleatorio consistente en observar el número de coches que están parados en un semáforo en rojo. El espacio muestral es:

$$X = 0, 1, 2, \ldots$$

La variable aleatoria $X \equiv$ número de coches toma un número contable e infinito de valores diferentes. Es decir, aunque en la práctica es imposible que haya infinitos coches parados en un semáforo, como el número máximo de coches es un valor elevado sin que esté claramente definido cual es, a la hora de formular un modelo matemático es más parsimonioso asumir que X no está acotada y que por tanto puede tomar infinitos valores.

Ejemplo 4. Si se define el experimento $T \equiv$ tiempo de espera en el semáforo, sabiendo que como máximo está en rojo un tiempo T_{max}, el espacio muestral es:

$$0 \leq T \leq T_{max}$$

Se trata de una variable continua y acotada. Toma un número infinito no numerable de valores, lo cual quiere decir que no es posible contar cuales son los valores que puede tomar T.

Las variables aleatorias, además de por el conjunto de valores que pueden tomar, se caracterizan por su función de distribución. La función de distribución, en ocasiones denominada función de distribución acumulada, asigna probabilidades a los valores de una variable aleatoria. En concreto, se define del siguiente modo:

> *Función de distribución* La *función de distribución* de una variable aleatoria X, indicada por $F(x)$, es una función con dominio la recta real e imagen el intervalo $[0, 1]$, tal que $F(x) = P(X \leq x)$.

Es decir, $F(x)$ es la probabilidad de que la variable X tome el valor x o algún valor menor de x. Sea $h > 0$, las propiedades más importantes de la función de distribución son:

1. $\lim_{x \to -\infty} F(x) = 0$.
2. $\lim_{x \to \infty} F(x) = 1$.
3. $F(x + h) - F(x) = P(x < X \leq x + h)$.
4. $F(x) \leq F(x + h)$.

Ejemplo 5. Sea T el tiempo en minutos que tarda una persona en ser atendido cuando acude a una ventanilla. Puede hipotetizarse que la función de distribución de T es:

$$F(t) = 1 - \exp(-t) \ .$$

Por ejemplo, la probabilidad de tener que esperar dos minutos o menos es:

$$\begin{aligned} F(2) &= P(T \leq 2) \\ &= 1 - \exp(-2) = 0,86 \ , \end{aligned}$$

y la probabilidad de tener que esperar más de tres minutos es $1 - F(3) = \exp(-3) = 0{,}05$. La figura 1 muestra la representación gráfica de $F(t)$ en función de t. Puede verse que se trata de una función creciente que toma valores entre 0 y 1.

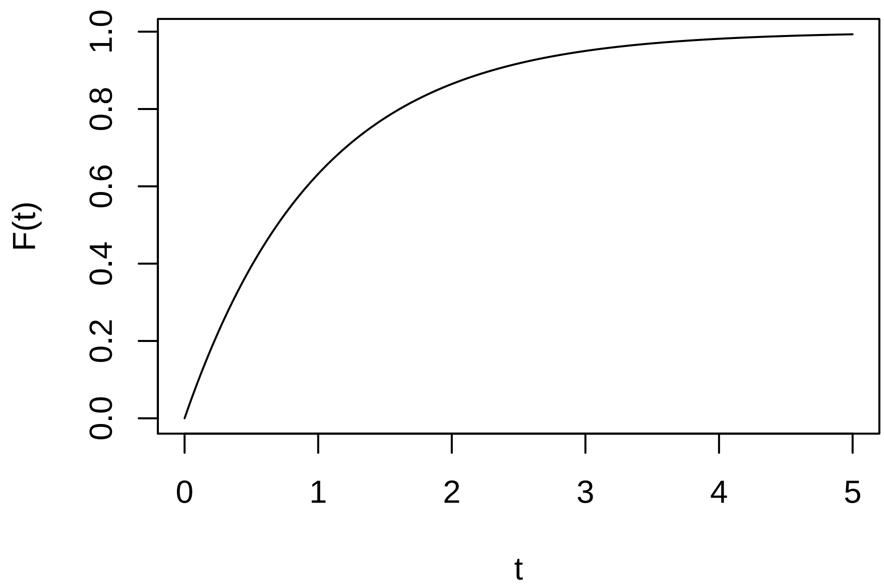

Figura 1: Probabilidad de esperar t minutos o menos

1.1. Variable aleatoria discreta

Una *variable aleatoria discreta* es aquella que toma un número contable de valores, que puede ser finito o infinito.

En los ejemplos 1 y 2 se han visto variables aleatorias discreta con un número finito de valores. En el ejemplo 3 la variable aleatoria también es discreta porque únicamente puede tomar valores enteros, sin embargo el espacio muestral es infinito dado que el número de valores de la variable no está acotado.

Función de probabilidad. Sea X una variable aleatoria discreta. La *función de probabilidad* de X, denominada $f(x)$, indica la probabilidad de que X tome el valor x. Es decir $f(x) = P(X = x)$.

Supongamos que los posibles valores de una variable discreta X se denominan $x_1, x_2, \ldots$, siendo $x_1 < x_2 < \ldots$. Entonces, la función de distribución de X se define como la probabilidad de obtener en una muestra el valor x_i o cualquiera de los valores más pequeños que x_i:

$$F(x_i) = P(X \leq x_i) = \sum_{j=1}^{i} f(X = x_j) \ .$$

Ejemplo 6. Sea la variable aleatoria $X \equiv$ número de personas que han contraído la gripe en el mes de diciembre en toda España. Se trata de una variable discreta con valores $X = 0, 1, 2, \ldots$. Obviamente en la realidad existe un máximo para el rango de valores que puede tomar X, que es el tamaño de la población de unos 47 millones de personas. En situaciones como esta, de cara a la modelización matemática es común asumir que no existe un máximo y que la variable X está definida en un espacio muestral infinito numerable. Además, supongamos que un investigador plantea la hipótesis de que la función de probabilidad de X viene dada por la expresión:

$$f(x) = \frac{2^x}{x!} \exp(-2).$$

Entonces, de acuerdo con la hipótesis del investigador, la probabilidad de que tres personas contraigan gripe es:

$$f(3) = \frac{2^3}{3!} \exp(-2) = \frac{8}{6} \exp(-2) \approx 0,18.$$

La probabilidad de que haya tres o menos personas que contraigan la gripe es

$$\begin{aligned} P(X \ \leq 3) &= F(3) \\ &= f(0) + f(1) + f(2) + f(3) \\ &\approx 0,86. \end{aligned}$$

En consecuencia, la probabilidad de que haya al menos cuatro personas que contraen la gripe es

$$P(X \geq 4) = 1 - P(X \leq 3) = 1 - F(3) \approx 0,14.$$

1.2. Variable aleatoria continua

Las variables aleatorias continuas toman un número infinito no numerable de valores. A diferencia de las variables discretas, los valores de una variable definida en un espacio muestral infinito no numerable no se pueden contar porque son números reales. Por ejemplo, sea una variable aleatoria con espacio muestral $X \geq 0$. Los valores de X no se pueden contar porque tienen infinitas cifras decimales.

La función de distribución, $F(x)$, se aplica tanto a variables discretas como continuas. En ambos casos $F(x)$ es la probabilidad de que X tome el valor x o cualquier otro valor menor que x. Sin embargo, con una variable continua la función de distribución se define utilizando una integral en lugar de un sumatorio.

> *Función de distribución.* Sea X una variable aleatoria continua. La función de distribución de X se define del siguiente modo:
>
> $$F(x) = \int_{-\infty}^{x} f(t)\, dt$$
>
> donde $f()$ es la denominada *función de densidad* de X.

La función de probabilidad no existe para las variables continuas porque cuando una variable puede tomar infinitos valores distintos, la probabilidad de que la variable tome un valor concreto es 0. Con una variable continua, $f(x)$ no indica la probabilidad del suceso $X = x$; en lugar de esto, a $f(x)$ se le denomina función de densidad de probabilidad o función de densidad por abreviar.

Las funciones $F(x)$ y $f(x)$ se relacionan mediante derivación e integración. Hemos visto que $F(x)$ es la integral de $f(x)$; a la inversa, $f(x)$ es la derivada de $F(x)$:

$$f(x) = F'(x).$$

Desde un punto de vista matemático, la probabilidad de que X tome valores en un intervalo concreto de valores se calcula mediante la integral definida de $f(x)$ en dicho intervalo. En concreto, la probabilidad de que X tome un valor comprendido entre a y b viene dada por:

$$
\begin{aligned}
P(a < X \leq b) &= F(b) - F(a) \\
&= \int_a^b f(x)dx \ .
\end{aligned}
$$

Según se ha mencionado, si X es continua la probabilidad de que tome un valor concreto es 0 debido a que la variable puede tomar infinitos valores. Matemáticamente esto se justifica atendiendo a la siguiente propiedad del cálculo integral:

$$
P(X = x) = \int_x^x f(t)\, dt = 0 \ .
$$

Para que una función matemática pueda ser utilizada como una función de densidad sólo debe cumplir dos propiedades:

1. $f(x) \geq 0$ para cada valor x de la variable X.
2. $\int_{-\infty}^{\infty} f(x)\, dx \ = \ 1$.

Lo cual no excluye los casos en que $f(x) > 1$. De hecho, es habitual encontrar funciones de densidad que toman valores superiores a 1, lo cual recuerda que $f(x)$ no es la probabilidad de x.

Ejemplo 7. Sea X una variable aleatoria uniforme definida en el intervalo $[0, \frac{1}{4}]$. Obtenga la función de densidad y la función de distribución de X.

Al ser X uniforme, la función de densidad es constante, $f(x) = c$. Por tanto, $f(x)$ satisface la primera propiedad siempre y cuando c sea no negativa. Con respecto a la segunda propiedad, debe cumplirse que $\int_0^{1/4} f(x)\, dx = 1$. Resolvemos la integral definida:

$$
\begin{aligned}
\int_0^{1/4} f(x)\, dx &= \int_0^{1/4} c\, dx \\
&= c\int_0^{1/4} dx \\
&= c[x]_0^{1/4} \\
&= c\left(\frac{1}{4} - 0\right) \\
&= \frac{c}{4}.
\end{aligned}
$$

Para que se cumpla $\int_0^{1/4} f(x)\, dx = 1$ basta con fijar $c = 4$, con lo que la función de densidad es

$$
f(x) = 4.
$$

La función de distribución, $F(x)$, indica la probabilidad de encortar valores entre 0 y x, y se obtiene mediante la integral definida de $f(x)$ entre el valor inferior de la variable (0 en este ejemplo) y el valor x para el cual queremos calcular su probabilidad acumulada:

$$
\begin{aligned}
F(x) &= \int_0^x 4\, dt \\
&= 4\int_0^x dt \\
&= 4[t]_0^x \\
&= 4x.
\end{aligned}
$$

Sea $T > 0$ una variable aleatoria con función de densidad

$$f(t) = \exp(-t)$$

La figura 2 muestra la representación gráfica de $f(t)$ para valores de T entre 0 y 1,5.

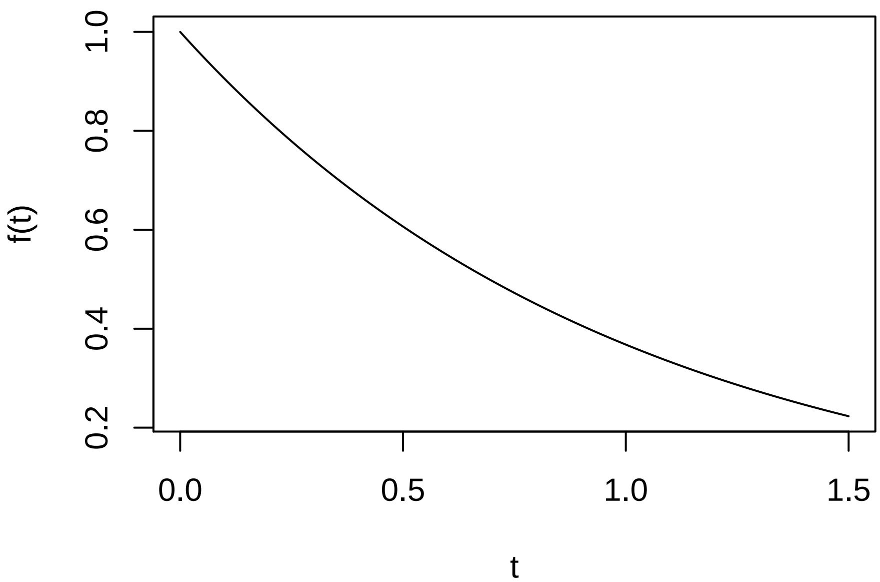

Figura 2: Representación gráfica de f(t) = exp(-t)

La función de distribución de T se obtiene a partir de $f(t)$ mediante integración:

$$\begin{aligned} F(t) &= \int_0^t \exp(-x)\,dx \\ &= (-\exp(-x))|_0^t \\ &= (-\exp(-t)) - (-\exp(0)) \\ &= 1 - \exp(-t). \end{aligned}$$

Ejemplo 8. Queremos saber si la siguiente función es una función de densidad:

$$f(x) = 2\left(1 - \frac{x^2}{25}\right), \text{ donde -5} \leq x \leq 5$$

Para resolver el problema podemos comenzar con la representación gráfica de la función de densidad, que aparece en la figura 3.

La función $f(x)$ es positiva en el intervalo (-5, 5), por lo que se satisface la primera propiedad. Sin embargo, su integral definida entre -5 y 5 no es igual a 1, en concreto:

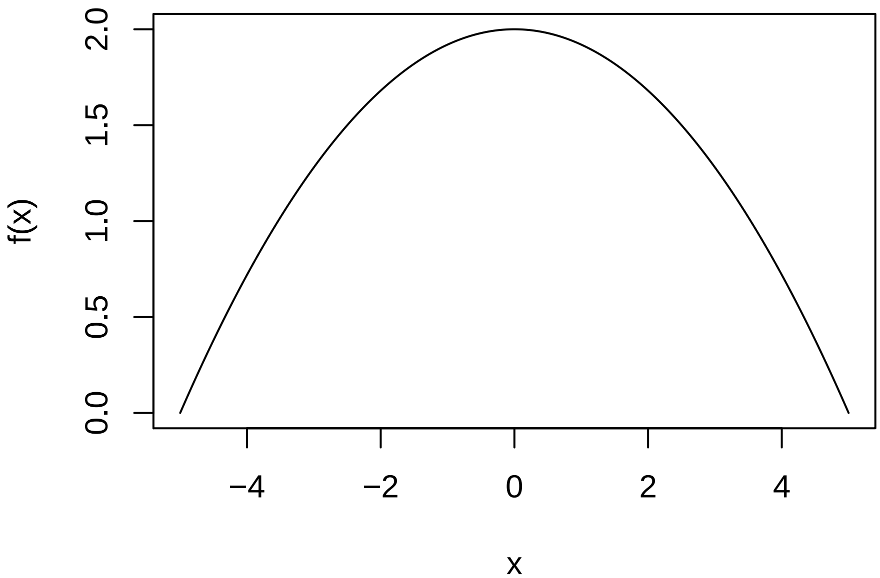

Figura 3: Función de densidad

$$
\begin{aligned}
\int_{-5}^{5} f(x)dx &= 2\int_{-5}^{5}\left(1-\frac{x^2}{25}\right)\,dx \\
&= 2\left(x-\frac{x^3}{75}\right)\Big|_{-5}^{5} \\
&= 2\left(\left[5-\frac{125}{75}\right]-\left[-5+\frac{125}{75}\right]\right) \\
&= \frac{1000}{75} \\
&= \frac{40}{3}
\end{aligned}
$$

Con lo cual la segunda propiedad no se cumple y $f(x)$ no es una función de densidad. Sin embargo, podemos definir la función

$$g(x) = \frac{3}{40}f(x)$$

que si es una función de densidad porque

$$\int_{-5}^{5} g(x)dx = \frac{3}{40}\int_{-5}^{5} f(x) = \frac{3}{40}\frac{40}{3} = 1.$$

La constante $3/40$ por la que se ha multiplicado $f(x)$ para obtener en una función de densidad se denomina en estadística *constante de integración*.

2. Valores esperados y momentos de una distribución

La función de distribución, $F(x)$, contiene toda la información acerca de X. Sin embargo, en ocasiones resulta excesivamente complejo describir una variable aleatoria utilizando su distribución. Por ello, se utilizan los denominados momentos para resumir la información contenida en una distribución y enfatizar determinados aspectos relevantes.

El momento más sencillo es el valor esperado de la variable aleatoria. Se denomina *valor esperado* de X y se indica por $E(X)$ a la siguiente expresión:

$$E(X) = \begin{cases} \sum_{i=1}^{I} x_i \ f(x_i), & \text{si } X \text{ es discreta, donde } I \text{ es el número de valores del espacio muestral} \\ \int_{-\infty}^{\infty} x \ f(x) dx, & \text{si } X \text{ es continua} \end{cases}$$

El valor esperado, $E(X)$, se denomina también *media poblacional* de X y se indica por la letra μ.

Ejemplo 9. Sea $X \in (0,1)$ una variable aleatoria continua con función de densidad

$$f(x) = 2x.$$

Su valor esperado es

$$\begin{aligned} E(X) &= \int_0^1 x f(x)\, dx \\ &= \int_0^1 2x^2\, dx \\ &= \left. \frac{2x^3}{3} \right|_0^1 \\ &= \frac{2}{3} \end{aligned}$$

El *momento de orden* r de una variable aleatoria X se define:

$$E(X^r) = \begin{cases} \sum_{i=1}^{n} x_i^r \ f(x), & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{\infty} x^r \ f(x)\, dx, & \text{si } X \text{ es continua} \end{cases}$$

Por ejemplo, los tres primeros momentos de una variable continua son:

$$\begin{aligned} \text{Valor esperado o media poblacional: } & \int_{-\infty}^{\infty} x \ f(x)\, dx \\ \text{Media de los valores al cuadrado: } & \int_{-\infty}^{\infty} x^2 \ f(x)\, dx \\ \text{Media de los valores al cubo: } & \int_{-\infty}^{\infty} x^3 \ f(x)\, dx \end{aligned}$$

También se utilizan los *momentos con respecto a la media*, que se definen del siguiente modo:

$$E((X-\mu)^r) = \begin{cases} \sum_{i=1}^{n} (x_i - \mu)^r \ f(x), & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{\infty} (x-\mu)^r f(x)\, dx, & \text{si } X \text{ es continua} \end{cases}$$

donde μ, es el valor esperado de la variable aleatoria: $\mu = E(X)$.

Los momentos con respecto a la media más utilizados son:

- La *varianza* de la variable, también indicado $Var(X)$ o σ^2, es el momento de orden dos con respecto a la media:

$$Var(X) = E((X-\mu)^2)\ .$$

  La varianza también puede calcularse mediante la expresión $Var(X) = E(X^2) - \mu^2$, que es equivalente a la anterior. La raiz cuadrada de la varianza se denomina *desviación típica* y se indica mediante σ.
- El momento de orden tres, que es un indicador de la *simetría* de la distribución: $E((X-\mu)^3)$.
- El momento de orden cuatro, que indica la *curtosis* o apuntamiento: $E((X-\mu)^4)$.

Ejemplo 10. Sea la función de densidad:

$$g(x) = \frac{3}{20}\left(1 - \frac{x^2}{25}\right), \quad -5 \leq x \leq 5$$

Su valor esperado es:

$$\begin{aligned}
E(X) &= \frac{3}{20}\int_{-5}^{5} x\left(1-\frac{x^2}{25}\right)dx \\
&= \frac{3}{20}\int_{-5}^{5}\left(x - \frac{x^3}{25}\right)dx \\
&= \frac{3}{20}\left(\frac{x^2}{2} - \frac{x^4}{100}\right)\Bigg|_{-5}^{5} \\
&= \frac{3}{20}\left(\frac{25}{2} - \frac{625}{100} - \frac{25}{2} + \frac{625}{100}\right) \\
&= 0
\end{aligned}$$

El momento de orden dos, α_2, es:

$$\begin{aligned}
E(X^2) &= \frac{3}{20}\int_{-5}^{5} x^2\left(1-\frac{x^2}{25}\right)dx \\
&= \frac{3}{20}\int_{-5}^{5}\left(x^2 - \frac{x^4}{25}\right)dx \\
&= \frac{3}{20}\left(\frac{x^3}{3} - \frac{x^5}{125}\right)\Bigg|_{-5}^{5} \\
&= \frac{3}{20}\left(\frac{125}{3} - \frac{3125}{125} + \frac{125}{3} - \frac{3125}{125}\right) \\
&= 5
\end{aligned}$$

Combinando ambas expresiones obtenemos la varianza de X:

$$Var(X) = E(X^2) - E(X)^2 = 5.$$

Ejemplo 11. Sea la distribución:

$$f(x) = 2(1-x), \quad 0 < x < 1$$

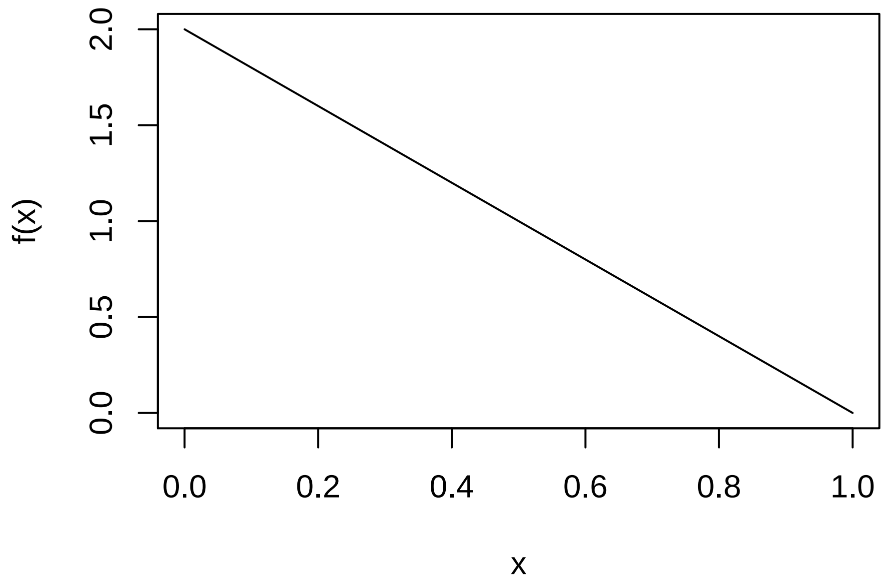

Figura 4: Representación gráfica de f(x) = 2(1-x)

Su representación gráfica aparece en la figura 4.

La media de esta distribución es:

$$\begin{aligned} E(X) &= 2\int_0^1 (x - x^2)dx \\ &= 2\left(\frac{x^2}{2} - \frac{x^3}{3}\right)\Big|_0^1 \\ &= \frac{1}{3}\ . \end{aligned}$$

El momento de orden dos es:

$$\begin{aligned} E(X^2) &= 2\int_0^1 (x^2 - x^3)dx \\ &= 2\left(\frac{x^3}{3} - \frac{x^4}{4}\right)\Big|_0^1 \\ &= \frac{1}{6}\ . \end{aligned}$$

Por tanto, la varianza de X es

$$Var(X) = \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{18}.$$

La asimetría de la distribución es:

$$
\begin{aligned}
E((X-E(X))^3) &= 2\int_0^1 \left(x-\frac{1}{3}\right)^3 (1-x)dx \\
&= 2\left(-\frac{x^5}{5}+\frac{x^4}{2}-\frac{4x^3}{9}+\frac{5x^2}{27}-\frac{x}{27}\right)\Big|_0^1 \\
&= \frac{1}{135}\ .
\end{aligned}
$$

3. Muestreo

3.1. Muestreo aleatorio simple

Según se ha visto, el espacio muestral Ω es el conjunto de resultados posibles en un experimento aleatorio, y una variable aleatoria X asigna a cada elemento de Ω un valor real. Se denomina *tomar una muestra* o *realizar* un experimento al proceso consiste en observar el valor de X en un determinado experimento. El valor observado de la variable X suele indicarse con la misma letra en minúscula que el nombre de la variable, x, y se le denomina *realización* de X.

Normalmente una muestra consta de varias observaciones y se indica por n al tamaño muestral. Cada una de las observaciones es la realización de una variable aleatoria: $X_1, X_2, \ldots, X_n$. En estadística, por motivos de simplicidad matemática, es habitual asumir que la muestra es aleatoria simple (m.a.s), es decir que las variables son independientes y están igualmente distribuidas.

Una *muestra aleatoria simple* es aquella que cumple dos condiciones:

1. Las variables $X_1, \ldots, X_n$ son independientes.
2. Las variables $X_1, \ldots, X_n$ tienen la misma función de probabilidad o densidad $f(x)$.

Como consecuencia de estas dos condiciones, la función de probabilidad de la muestra es el producto de la probabilidad de cada una de las observaciones. Supongamos que el vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ representa la muestra; su función de probabilidad o densidad es el producto:

$$
\begin{aligned}
f(\boldsymbol{x}) &= f(x_1, \ldots, x_n) \\
&= f(x_1)\cdots f(x_n).
\end{aligned}
$$

Ejemplo 12. Supongamos que la variable aleatoria X, definida intervalo $(0, \infty)$, indica el número de horas de duración continuada de un producto hasta que falla. La función de densidad de X es

$$
f(x) = x\exp(-x)\ ,
$$

Si tomamos una muestra aleatoria de tamaño dos, $\boldsymbol{x} = (x_1, x_2)$, su función de densidad es

$$
\begin{aligned}
f(\mathbf{x}) &= f(x_1)f(x_2) \\
&= (x_1\exp(-x_1)) \times (x_2\exp(-x_2)) \\
&= x_1x_2\exp(-x_1-x_2).
\end{aligned}
$$

De modo general, si la muestra contiene un número n de observaciones, entonces viene descrita por el vector $\mathbf{x}' = (x_1, \ldots, x_n)$, y la función de densidad de la muestra es

$$
\begin{aligned}
f(\mathbf{x}) &= f(x_1)\cdots f(x_n) \\
&= x_1\cdots x_n \exp(-x_1-\cdots-x_n) \\
&= \prod_{i=1}^{n} x_i \exp\left(-\sum_{i=1}^{n} x_i\right).
\end{aligned}
$$

Además, si tenemos en cuenta que la media muestral es $\overline{X} = \sum_{i=1}^{n} x_i/n$, podemos escribir la función de probabilidad de la muestra de un modo más compacto

$$
f(\mathbf{x}) = \prod_{i=1}^{n} x_i \exp\left(-n\overline{X}\right).
$$

Por concretar con unos datos, supongamos que se ha medido la duración de tres productos y se obtiene el resultado $\boldsymbol{x} = (5,\, 2,\, 8)$. Entonces, la función de densidad de la muestra es

$$
f(5,\, 2,\, 8) = 80 \exp(-15).
$$

Ejemplo 13. Supongamos que π es la proporción de cerezas sanas en una cesta y la variable X toma el valor 0 cuando una cereza no está sana y el valor 1 cuando si lo está. La función de probabilidad de X es

$$
f(x) = \pi^x(1-\pi)^{1-x}.
$$

Es fácil ver que la probabilidad de obtener una cereza sana es $f(X=1) = \pi^1(1-\pi)^{1-1} = \pi$, y la probabilidad de obtener una que no esté sana es $f(X=0) = \pi^0(1-\pi)^{1-0} = 1-\pi$. Supongamos que una persona toma dos cerezas y el resultado se indica mediante $\mathbf{x}' = (x_1, x_2)$. Por simplificar asumimos que el muestreo es con reposición y por tanto la proporción de cerezas sanas, π, no cambia de una extracción a otra. En consecuencia, la función de probabilidad de $\boldsymbol{x}$ es

$$
\begin{aligned}
f(\boldsymbol{x}) &= f(x_1)f(x_2) \\
&= (\pi^{x_1}(1-\pi)^{1-x_1})(\pi^{x_2}(1-\pi)^{1-x_2}) \\
&= \pi^{x_1+x_2}(1-\pi)^{2-x_1-x_2}.
\end{aligned}
$$

De modo general, si $\mathbf{x}' = (x_1, \ldots, x_n)$ es un vector de n observaciones en el cual hay $s = \sum_i x_i$ cerezas sanas, la función de probabilidad de $\mathbf{x}$ es

$$
f(\boldsymbol{x}) = \pi^s(1-\pi)^{n-s}.
$$

3.2. Muestreo aleatorio

El muestreo aleatorio se utiliza con frecuencia en el contexto de la psicología matemática del aprendizaje. El muestreo aleatorio asume que existe una función de probabilidad asociada a cada una de las realizaciones de la muestra, pero a diferencia del muestreo aleatorio simple, dicha función no es igual para todos los elementos de la muestra.

Ejemplo 14. Supongamos que un sujeto experimental tiene que realizar una tarea repetidamente. La variable X indica el resultado de dicha tarea, que puede ser correcto ($X = 1$) o incorrecto ($X = 0$). Asumimos que existe una probabilidad de acierto la primera vez que se realiza la tarea que es igual a π. Es decir, si X_1 indica el resultado de la primera realización, su función de probabilidad es

$$f(x_1) = \pi^{x_1}(1-\pi)^{1-x_1}$$

Es fácil ver que la probabilidad de acierto es $f(1) = \pi$ y la probabilidad de fallo $f(0) = 1 - \pi$.

Supongamos que el modelo asume que existe aprendizaje. En concreto, cuando la tarea se realiza correctamente se produce aprendizaje y aumenta la probabilidad de acertar en los sucesivos ensayos. Cuando se falla, no se produce aprendizaje. Entonces, puede hipotetizarse que la probabilidad de realización correcta en el ensayo k es

$$\pi^{1-y_k}$$

donde y_k es la proporción de aciertos antes de llegar al ensayo k:

$$y_k = \frac{\sum_{j=1}^{k-1} x_j}{k-1}$$

La figura 5 muestra como cambia la probabilidad de acierto en función de la proporción de aciertos en los ensayos previos asumiendo que el nivel de partida es $\pi = 0,2$.

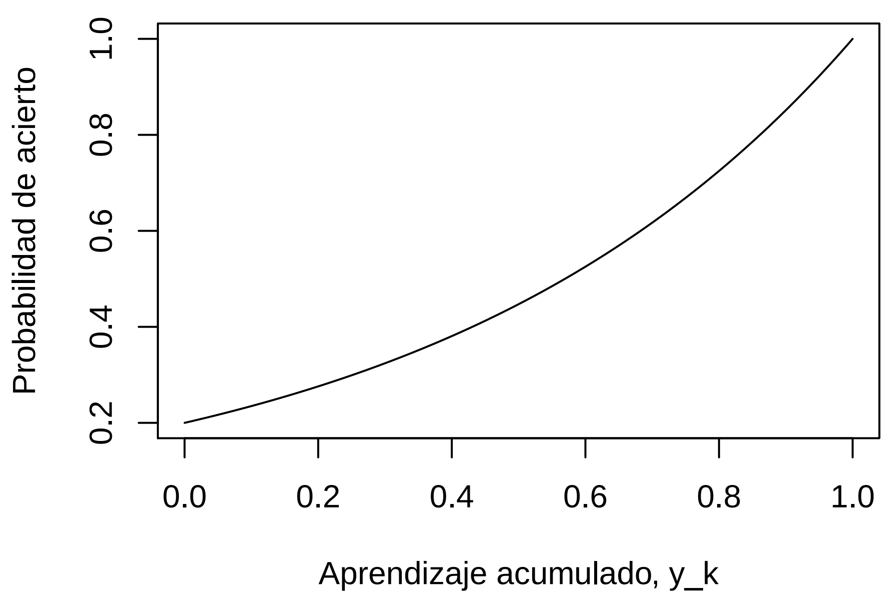

Figura 5: Probabilidad de acierto en función del aprendizaje acumulado

Por tanto, en este modelo la probabilidad de acertar en una tarea cambia en función de los resultados que el sujeto haya obtenido en las tareas anteriores. En concreto, la probabilidad de una muestra de tres observaciones se define

$$\begin{aligned} f(\boldsymbol{x}) &= f(x_1)f(x_2)f(x_3) \\ &= \left[\pi^{x_1}(1-\pi)^{1-x_1}\right] \times \left[\pi^{(1-y_2)x_2}(1-\pi^{1-y_2})^{1-x_2}\right] \times \left[\pi^{(1-y_3)x_3}(1-\pi^{1-y_3})^{1-x_3}\right] \end{aligned}$$

4. Prácticas en R

El lenguaje R permite conocer de forma aproximada las propiedades de una función de distribución teórica. En primer lugar veremos como representar gráficamente las funciones de densidad y distribución para obtener una impresión de la forma de estas funciones y de en qué valores de la variable se concentra una mayor probabilidad. En segundo lugar, veremos cómo maximizar funciones en R, lo que permite encontrar la moda de las distribuciones y -según veremos en capítulos posteriores- estimar parámetros. Finalmente calcularemos de forma numérica los momentos de una distribución.

4.1. Representación gráfica de funciones

Para representar gráficamente una función, en primer lugar tenemos que definir el conjunto de valores que forman el eje de ordenadas y lo guardamos en un vector. A continuación obtenemos el valor de la función para cada uno de dichos valores y lo guardamos en un segundo vector. Después le pasamos ambos valores a la función `plot`. Por ejemplo, supongamos que vamos a representar la siguiente función

$$f(x) = 6x(1-x)\,,$$

definida en el intervalo (0, 1). El siguiente código R muestra cómo hacer la representación gráfica. La primera línea de código utiliza el comando `seq` para definir el vector de valores $x = 0,\ 0{,}01,\ 0{,}02,\ \ldots,\ 1$. La segunda línea calcula la función $f(x) = 6x(1-x)$ y guarda sus valores en el vector f. La tercera línea contiene el comando `plot` para representar el gráfico utilizando los pares de puntos $(x_i,\ f_i)$. El resultado aparece en la figura 6.

```
x <- seq(0, 1, by=0.01)
f <- 6*x*(1-x)
plot(x, f, type="l", lwd=2, ylab="", col="red")
```

A continuación veremos cómo representar varias funciones en los mismos ejes de coordenadas. Esto ocurre por ejemplo cuando queremos representar dos funciones de densidad distintas o una función de densidad y su correspondiente función de distribución. Vamos a hacer una gráfica con la misma función $f(x)$ vista en el ejemplo anterior y además la función

$$g(x) = 18x(1-x)^4$$

Para ello, en primer lugar definimos los valores de x y calculamos los valores de f y g. A continuación hay que representar $f(x)$ mediante el comando `plot`, y se utiliza el comando `lines` para añadir líneas a un gráfico ya existente. Por último, utilizando `legend` se añade una leyenda a la figura para especificar cual es la línea correspondiente a cada función. En la figura 7 puede verse el resultado.

```
x <- seq(0, 1, by=0.01)
f <- 6*x*(1-x)
g <- 18*x*(1-x)^4
plot(x, f, type="l", lwd=2, ylab="", ylim=c(0,1.6), col="red")
lines(x, g, lwd=2, col="blue")
legend(0.75, 1.6, c("f(x)", "g(x)"), lty=c(1,1), lwd=2, col=c("red", "blue"))
```

4.2. Maximización de funciones

En matemáticas se denomina optimizar una función a buscar su valor máximo o mínimo. El lenguaje R tiene incorporadas diversas funciones para realizar esta tarea, además de existir paquetes adicionales de R destinados a aumentar sus capacidades de optimización de funciones. En este apartado veremos cómo utilizar

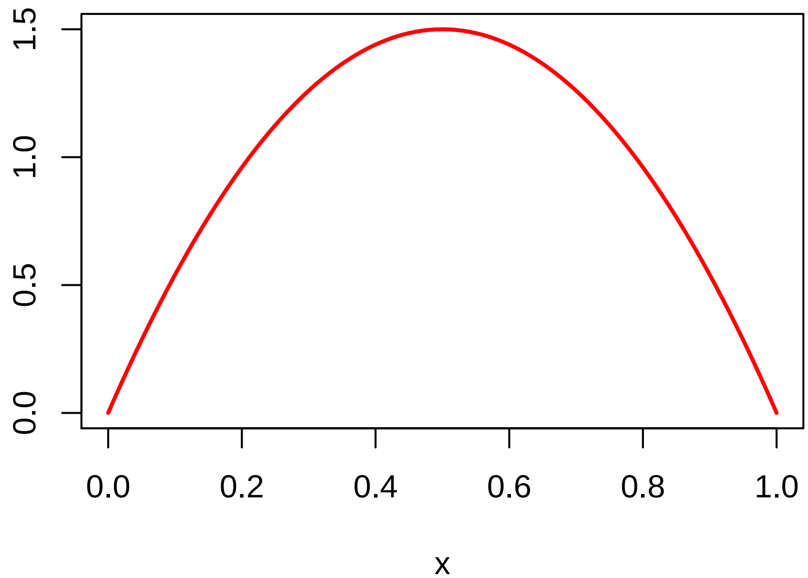

Figura 6: Representación gráfica de f(x) = 6x(1-x)

la función más sencilla de optimización en R, denominada `optimize`. Vamos a aplicar esta función para buscar el máximo de la función de densidad $g(x) = 18x(1-x)^4$ representada gráficamente en el apartado anterior.

Según se ha visto en cursos anteriores, se denomina moda de una distribución al valor de la variable X para el cual su función de probabilidad o densidad alcanza un máximo. La moda podemos obtenerla acudiendo a los procedimientos del cálculo matemático. Para ello, partimos de la observación de que el máximo de $g(x)$ y el máximo de la función $h(x) = \log g(x)$ están en el mismo punto; esto se debe a que el logaritmo es una función monótona, por lo que si $x_1 > x_2$ entonces $\log(x_1) > \log(x_2)$. Por tanto:

$$h(x) = \log g(x) = \log 18 + \log x + 4\log(1-x)\,.$$

Gracias a las propiedades de la función logaritmo, podemos trabajar con $h(x)$ en lugar de con $g(x)$ y ganar así en sencillez. En el máximo de una función se cumple que su derivada es cero, por lo que obtenemos la derivada de $h(x)$, que tiene la forma

$$h'(x) = \frac{1}{x} - \frac{4}{1-x}\,.$$

Entonces buscamos el valor de x para el que se cumple que $h'(x) = 0$. Dicho valor es:

$$\begin{aligned} \frac{1}{x} - \frac{4}{1-x} &= 0 \\ 1-x &= 4x \\ x &= \frac{1}{5}. \end{aligned}$$

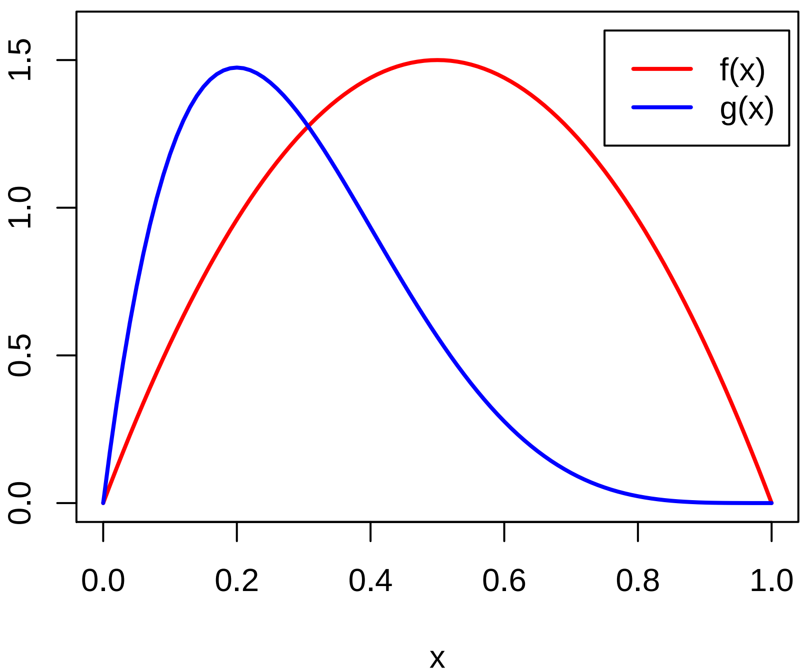

Figura 7: Representación de dos funciones en una misma gráfica

Hemos demostrado matemáticamente que la moda es $1/5$. Ahora pondremos a prueba a la función `optimize` para obtener la moda con R. La función `optimize` necesita dos argumentos de entrada: la función a maximizar y el intervalo de valores en el que tiene que buscar el máximo. Además es necesario indicarle a `optimize` que busque el máximo de la función, porque por defecto lo que hace es buscar el mínimo de la función que le hayamos pasado. El siguiente código R ilustra el procedimiento, la primera línea de código define la función $g(x)$ y en la segunda se llama a `optimize` para que guarde el resultado en el objeto `fit`. En la llamada a *optimize* le pasamos la función $g(x)$, le decimos que el intervalo de valores de x donde tiene que buscar el máximo de $g(x)$ es (0, 1), y le decimos que busque el máximo de la función. En la tercera linea de código se llama a `print` para que muestre los elementos contenidos en el objeto `fit`, que son el valor de x donde se encuentra el máximo y el valor de $g(x)$ en dicho punto.

```
g <- function(x) 18 * x * (1-x)^4
fit <- optimize(g, c(0, 1), maximum=TRUE)
print(fit)
```

```
## $maximum
## [1] 0.2000195
##
## $objective
## [1] 1.47456
```

Al ejecutar este código puede advertirse que el resultado de R no es exacto, ya que indica que el valor de x es 0,2000195 cuando antes habíamos comprobado que el valor correcto es $1/5 = 0{,}2$. Esto es algo habitual cuando se trabaja con métodos numéricos, que por definición proporcionan solo resultados aproximados.

4.3. Cálculo de los momentos de una distribución

El lenguaje R incorpora funciones para calcular integrales definidas de forma numérica. El lenguaje R está orientado al cálculo numérico más que al simbólico. El cálculo simbólico consiste en desarrollar fórmulas matemáticas, por ejemplo obteniendo la función derivada de otra función, la integral, simplificar expresiones, etc. Para realizar estas tareas es mejor emplear otro tipo de programas informáticos. R en cambio tiene buenas capacidades para calcular el resultado numérico de una expresión, por ejemplo una integral definida, lo que puede aplicarse a la obtención de los momentos de una distribución. Esto quiere decir que es adecuado para calcular el resultado que proporcionan las fórmulas matemáticas, pero no para desarrollar estas

La función `integrate` permite evaluar numéricamente integrales definidas. A modo de ejemplo, supongamos que queremos calcular en R qué valor toma la integral definida de la función $f(x) = 6x(1-x)$ entre 0 y 1. Podemos obtenerlo mediante las tres líneas de código que aparecen a continuación, la primera línea define la función a integrar, la segunda calcula numéricamente la integral definida llamando al comando `optimy` la tercera muestra el resultado utilizando el comando `print`.

```
f1 <- function(x) {6*x*(1-x)}
integral <- integrate(f1, lower = 0, upper = 1)
print(integral)
```

```
## 1 with absolute error < 1.1e-14
```

Desde el punto de vista de la teoría de la probabilidad, dado que $f(x)$ toma valores no negativos en el rango $(0, 1)$ y su integral es 1, entonces es una función de densidad. Ahora podemos plantearnos obtener su valor esperado, que no es más que el resultado de la integral

$$E(x) = \int_0^1 x f(x)\, dx = \int_0^1 6x^2(1-x)\, dx\,.$$

Utilizaremos para ello el siguiente código

```
xfx <- function(x) {6*x^2*(1-x)}
Ex <- integrate(xfx, lower = 0, upper = 1)
print(Ex)
```

```
## 0.5 with absolute error < 5.6e-15
```

Al ejecutar el código vemos que el valor esperado es 0,5, como no puede ser de otra manera observando la forma de $f(x)$ en la figura 6. La función `integrate` proporciona un objeto de R, que es un conjunto de elementos agrupados bajo un nombre común. Si quisiéramos saber qué elementos contiene el objeto `Ex` que hemos obtenido como resultado de la ejecución del código anterior, escribiríamos `names(Ex)`. El lenguaje R nos informa entonces de que el objeto `Ex` contiene los campos `value`, `abs.error`, `subdivisions`, `message` y `call`. Para acceder a estos campos utilizamos el signo `$`. El campo más importante es `value`, que contiene el resultado numérico de la integral.

Continuando con el ejemplo, supongamos que queremos obtener la varianza de esta distribución. Podemos hacerlo obteniendo el valor esperado de X al cuadrado, $E(X^2)$ y restándole la media poblacional al cuadrado para implementar la fórmula $\sigma^2 = E(X^2) - E(X)^2$. Los comandos de R son:

```
x2fx <- function(x) {6*x^3*(1-x)}
Ex2 <- integrate(x2fx, lower = 0, upper = 1)
VarX <- Ex2$value - Ex$value^2
print(VarX)
```

```
## [1] 0.05
```

Además de para calcular momentos, la función `integrate` nos permite calcular la probabilidad de un intervalo integrando la función de densidad entre los dos extremos de dicho intervalo. Por ejemplo, queremos saber cuanto vale la probabilidad de que la variable X tome un valor superior a 0,75. Matemáticamente el problema sería entonces obtener la integral de la función de densidad entre 0,75 y el límite superior de la variable aleatoria. Es decir

$$P(X \geq 0,75) = \int_{0,75}^{1} 6x(1-x)\,dx\,.$$

Para resolverlo utilizamos el código

```
Pr <- integrate(f1, lower = 0.75, upper = 1)
print(Pr)
```

```
## 0.15625 with absolute error < 1.7e-15
```

Por como se indicaba anteriormente, a diferencia de otros sistemas informáticos de cálculo simbólico, el lenguaje R solo realiza cálculo numérico y sus resultados están sujetos a errores de aproximación. Por ejemplo, el valor esperado de la distribución $f(x) = 2(1-x)$ es exactamente

$$E(x) = 2\int_{0}^{1} 2(x-x^2)\,dx = \frac{1}{3}\,.$$

Vamos a replicar este resultado en R con el código

```
fx <- function(x) {2*(x-x^2)}
Ex <- integrate(fx, lower = 0, upper = 1)
print(Ex)
```

```
## 0.3333333 with absolute error < 3.7e-15
```

La ejecución de R da como resultado $E(X) = 0,33333$ lo cual es matemáticamente falso porque dicha cantidad no es igual a 1/3. El problema se debe a que el computador necesariamente debe trabajar con un número finito de valores decimales para calcular los resultados, y solo puede proporcionar una aproximación a un resultado exacto que en realidad tiene infinitos decimales.

5. Ejercicios

Ejercicio 1. Se introducen dos ratas en un laberinto con cuatro salidas. Sea X el número de ratas que salen por la primera salida. Asumiendo que cada rata puede salir por cualquier salida con igual probabilidad y que el comportamiento de cada una es independiente de la otra.

1. ¿Cuanto vale la probabilidad de los sucesos $X = 0$, $X = 1$ y $X = 2$?
2. Obtenga la media y varianza de X .

Ejercicio 2. Dada la siguiente función:

$$g(x) = 3x^2, \quad x \in (0, 1),$$

Responda a los siguientes apartados:

1. Compruebe si $g(x)$ es una función de densidad.
2. Obtenga la función de distribución $G(x)$.
3. Obtenga el valor esperado y varianza de X matemáticamente y mediante R.
4. Represente gráficamente $g(x)$ y $G(x)$.

Ejercicio 3. Sea X una variable aleatoria con rango $-1 \leq X \leq 1$ y función de densidad

$$f(x) = \frac{3}{4}(1 - x^2).$$

1. Represente $f(x)$ gráficamente .
2. Obtenga $E(X)$.
3. Obtenga $Var(X)$.
4. Obtenga la moda de $f(x)$ matemáticamente y mediante R.

Ejercicio 4. Sea X una variable aleatoria con rango $-1 \leq X \leq 1$ y función de densidad

$$f(x) = \frac{3}{4}(1 - x^2).$$

1. Represente $f(x)$ gráficamente .
2. Obtenga $E(X)$.
3. Obtenga $Var(X)$.

Ejercicio 5. Sea X una variable aleatoria con función de densidad:

$$f(x) = \frac{1 + \alpha x}{2}, \quad -1 \leq x \leq 1,$$

Obtenga $E(X)$ y $Var(X)$.

Ejercicio 6. Sea $X \geq 0$ una variable aleatoria con función de densidad

$$f(x) = \exp(-x)$$

1. Utilizando R, obtenga $P(X \leq 2)$.
2. Obtenga la función de densidad de una muestra aleatoria simple de tamaño n.
3. Sea $\overline{X}$, utilizando R obtenga $P(\overline{X} \leq 2)$ en una muestra de tamaño 4.

Ejercicio 7. Sea X una variable aleatoria con función de densidad:

$$f(x) = \frac{1 + \alpha x}{2}, \quad -1 \leq x \leq 1,$$

Obtenga $E(X)$ y $Var(X)$.

Ejercicio 8. La función de distribución de X es:

$$F(x) = 1 - \exp(-x/\alpha) \ ,$$

siendo $x \geq 0$.

1. Obtenga $f(x)$.
2. Obtenga $E(X)$.
3. Obtenga $f(\mathbf{x})$, siendo $\mathbf{x}$ una muestra aleatoria simple de tamaño n.