

Tutorial 05: Teorema Central del Límite.

Atención:

- Este documento pdf lleva adjuntos algunos de los ficheros de datos necesarios. Y está pensado para trabajar con él directamente en tu ordenador. Al usarlo en la pantalla, si es necesario, puedes aumentar alguna de las figuras para ver los detalles. Antes de imprimirlo, piensa si es necesario. Los árboles y nosotros te lo agradeceremos.
- Fecha: 19 de abril de 2017. Si este fichero tiene más de un año, puede resultar obsoleto. Busca si existe una versión más reciente.

Índice

1. La Distribución Binomial.	1
2. Zoo binomial	9
3. Variables aleatorias continuas. Funciones e integración con GeoGebra, Wiris y Wolfram Alpha.	14
4. La distribución normal.	33
5. La distribución normal en R.	38
6. Definición de funciones e integración en R	42
7. Ejercicios adicionales y soluciones.	49

1. La Distribución Binomial.

La distribución binomial (ver la Sección 5.1.2, pág. 129 del libro) es la primera de las grandes distribuciones clásicas, o distribuciones distinguidas, con nombre propio, que encontramos en el curso. No será, desde luego, la última. Después vendrán la normal, la t de Student, la de Poisson, y otras cuantas. Como veremos, una ventaja, al trabajar con R, es que el tratamiento de todas esas distribuciones es muy parecido. Lo que vamos a aprender aquí para la binomial nos servirá, con algunas modificaciones menores, para todas las distribuciones que vendrán detrás. Por contra, como también veremos, al usar una hoja de cálculo (como Calc), el trabajo con cada una de esas distribuciones es distinto y mucho menos cómodo.

1.1. Simulando un ejemplo básico en R.

Para que nos sirva de calentamiento, vamos a empezar usando R para ayudarnos con la lectura del Ejemplo 5.1.1 del libro (pág. 129). En ese ejemplo lanzamos un dado cinco veces, y nos preguntamos por la probabilidad de obtener exactamente dos seises en esas cinco tiradas. El primer paso es utilizar la librería `gtools`, que hemos visto en el Tutorial03, para construir la lista completa de resultados que pueden obtenerse al tirar cinco veces un dado. Esa lista la forman las permutaciones con repetición de seis elementos, tomados de cinco en cinco. La lista que obtenemos como respuesta se almacena en una matriz, que hemos llamado `tiradas`. Usaremos `head` y `tail` para ver el aspecto de esta matriz:

```

rm(list=ls())
library(gtools)
tiradas = permutations(n=6, r = 5, repeats.allowed = TRUE)
head(tiradas)

##      [,1] [,2] [,3] [,4] [,5]
## [1,]    1    1    1    1    1
## [2,]    1    1    1    1    2
## [3,]    1    1    1    1    3
## [4,]    1    1    1    1    4
## [5,]    1    1    1    1    5
## [6,]    1    1    1    1    6

tail(tiradas)

##      [,1] [,2] [,3] [,4] [,5]
## [7771,]    6    6    6    6    1
## [7772,]    6    6    6    6    2
## [7773,]    6    6    6    6    3
## [7774,]    6    6    6    6    4
## [7775,]    6    6    6    6    5
## [7776,]    6    6    6    6    6

```

Lo que hacemos ahora es usar los trucos que hemos aprendido en simulaciones de los tutoriales previos para localizar las partidas ganadoras, y calcular su frecuencia relativa como estimación de la probabilidad:

```

es6 = (tiradas==6)
partidasGanadoras = (rowSums(es6) == 2)
cuantasGanadoras = sum(partidasGanadoras)

(probabilidadGanar = cuantasGanadoras / nrow(tiradas))

## [1] 0.1608

```

¡Pero cuidado con malinterpretar lo que estamos haciendo! Aunque las técnicas son las mismas que las de las simulaciones, lo que hemos hecho aquí es construir el espacio muestral completo, y contar los casos en los que ocurre el suceso que nos interesa (la partida es ganadora). No es una simulación, es una enumeración. Y por lo tanto el resultado

1250 partidas ganadoras de un total de 7776 partidas,

tiene que ser exactamente el mismo que produce el cálculo con la distribución binomial y que, como vimos en el ejemplo del libro, es:

$$\frac{\binom{5}{2} 5^3}{6^5} = \frac{1250}{7776} = \frac{625}{3888} \approx 0.1608$$

1.1.1. La distribución binomial en R.

A continuación vamos a aprender a calcular directamente valores como estos en R. Para trabajar con la distribución binomial R nos ofrece básicamente cuatro funciones, que son

`dbinom, pbinom, qbinom, rbinom.`

Todas ellas, como se ve, comparten el final, el sufijo `binom`, pero cambian en la letra inicial, que puede ser `d`, `p`, `q` o `r`. Esto es un principio general en R. Cuando aprendamos a trabajar con la distribución normal veremos que el sufijo correspondiente es `norm`, y que hay cuatro funciones análogas, llamadas

`dnorm`, `pnorm`, `qnorm`, `rnorm`,

que cumplen cada una un papel similar, indicado por la letra inicial. Vamos por tanto, una por una, a entender lo que hacen estas cuatro funciones que acaban en `binom`.

1.2. La función `dbinom`.

La función `dbinom` es la función de densidad de una variable aleatoria binomial. Como hemos visto en la Ecuación 5.4 (pág. 134) del Capítulo 5 del libro, si esa variable es de tipo $B(n, p)$, entonces su función de densidad viene dada por:

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}.$$

Por ejemplo, si $n = 25$, $p = 1/3$ (con lo que $q = 2/3$) y queremos calcular (con $k = 6$):

$$P(X = 6) = \binom{25}{6} \cdot \left(\frac{1}{3}\right)^6 \cdot \left(\frac{2}{3}\right)^{25-6}.$$

entonces la función `dbinom` permite calcular este valor así:

```
dbinom(6, size= 25, prob=1/3)
```

```
## [1] 0.1096
```

Si estamos trabajando con la binomial $B(n, p)$, la función `dbinom` se puede aplicar al vector `0:n` de \mathbb{R} para obtener la tabla de densidad de probabilidad de esa distribución. Si, por ejemplo, estamos trabajando con una binomial $B(10, 1/5)$, esa tabla se obtiene con:

```
dbinom(0:10, size= 10, prob=1/5)
```

Ejercicio 1.

1. Encuentra esos valores de probabilidad, y busca el máximo entre ellos.
2. Usa la función `dbinom` para comprobar los cálculos del Ejemplo 5.1.1 del libro (pág. 129), que es el que hemos usado para abrir este tutorial. Recuerda que lo que hemos hecho en este tutorial no es una simulación, sino una enumeración, así que los dos resultados deben ser idénticos. No parecidos, sino idénticos.
3. Calcula la probabilidad de sacar exactamente 3 veces el 5 al lanzar un dado de 17 caras 11 veces.
4. En el Ejemplo 5.3.1 del libro (página 144), tenemos una variable X , de tipo $B(n, p)$ con $n = 21$, $p = \frac{1}{3}$ y queremos calcular el valor exacto de

$$P(5 \leq X \leq 9) = P(X = 5) + P(X = 6) + \dots + P(X = 9).$$

Usa la función `dbinom` para comprobar el resultado 0.7541 (4 cifras sig.) que aparece en el texto.

5. Usa también `dbinom` para calcular, de forma aproximada, los valores de probabilidad que aparecen en la Tabla 5.5 del libro (Ejemplo 5.1.4, pág. 135).

Solución en la página 56. □

1.3. La función `pbinom`

Si la función `dbinom` es la función de densidad de la distribución binomial, la función `pbinom` es su función de distribución (recuerda la definición de la Sección 4.4, pág. 111 del libro). Es decir, que `pbinom` nos proporciona, para cada uno de los valores

$$k = 0, 1, 2, \dots, n,$$

la probabilidad de que X tome un valor menor o igual que k :

$$F(X = k) = P(X \leq k).$$

Se trata de probabilidades acumuladas, las análogas teóricas de las frecuencias acumuladas. Por ejemplo, para la binomial $B(10, 1/5)$, la probabilidad de que X tome un valor igual o inferior a 3 es:

```
pbinom(3, size= 10, prob=1/5)
## [1] 0.8791
```

Si la función `dbinom` era muy fácil de entender, con la `pbinom` tenemos que empezar a tener un poco de cuidado. Especialmente, **porque la distribución binomial es discreta**. Para ilustrar lo que queremos decir, supongamos que, de nuevo con una binomial $B(10, 1/5)$, queremos ahora la probabilidad de que X tome un valor igual o superior a 3. Es decir,

$$P(X \geq 3)$$

R no tiene una función para calcular esto directamente, debemos usar `pbinom` y las propiedades de la Probabilidad. El suceso contrario de $X \geq 3$ es $X < 3$. Así que

$$P(X \geq 3) = 1 - P(X < 3)$$

y ahora necesitamos calcular

$$P(X < 3).$$

Y el peligro aquí es confundir las desigualdades estrictas (que son $<$ y $>$) con las que no lo son (es decir, \leq y \geq). **La función `pbinom` se define con \leq** . Así que tenemos que expresar la probabilidad que queremos usando \leq . Y en este caso, eso significa darse cuenta de que, *como X (de tipo $B(10, 1/5)$) es discreta*, sólo toma los valores

$$0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$$

y ningún otro valor. Es decir, que (y este es el **paso clave**)

$$P(X < 3) = P(X \leq 2).$$

Y esta última probabilidad es la que podemos calcular con

```
pbinom(2, size= 10, prob=1/5)
## [1] 0.6778
```

El punto clave, insistimos, es que para calcular $P(X < 3)$ hemos tenido que reescribirlo como $P(X \leq 2)$ (porque la binomial es discreta). Los siguientes ejercicios pretenden ayudarte a practicar estas ideas.

Ejercicio 2.

En una binomial $B(7, 1/4)$, calcula las siguientes probabilidades:

1. $P(X \leq 4)$.
2. $P(X < 3)$.
3. $P(X \geq 2)$.
4. $P(X > 3)$.
5. $P(2 \leq X \leq 4)$ (o, lo que es lo mismo, $P((X \geq 2) \cap (X \leq 4))$).
6. $P(2 < X < 4)$.
7. $P(2 \leq X < 4)$.
8. $P(2 < X \leq 4)$.

¿Echas de menos algún caso? Cuando estés convencido de tener una lista completa de preguntas sobre la binomial, escribe un fichero de instrucciones R en el que baste con introducir n , p , y el límite o límites de la desigualdad para obtener todas las probabilidades de una vez. Solución en la página 57. \square

La moraleja de todos estos ejercicios es la siguiente:

Nunca debemos olvidar que, en R, todas las funciones de distribución, que empiezan por `p`, como `pbinom` calculan la probabilidad usando \leq . A menudo diremos que usan la cola izquierda de la distribución.

Ejercicio 3.

Usa la función `barplot` para representar gráficamente los valores de `dbinom` y `pbinom` en la binomial $B(7, 1/4)$ (son dos gráficas distintas). Pero antes trata de imaginar el resultado (usa lápiz y papel para hacer un esbozo) y la relación entre ambas gráficas. Solución en la página 58. \square

1.4. La función `qbinom`

La función `qbinom` nos permite calcular los cuantiles de una variable aleatoria binomial. Hemos hablado de los cuantiles en la Sección 4.4 del libro, donde hemos dicho que eran los análogos teóricos de los cuartiles y percentiles de la Estadística Descriptiva. Un ejemplo ayudará a entenderlo mejor. Dada una variable binomial, de tipo $B(10, 0.3)$, ¿cuál es el primer valor (de 0 a 10) para el que la probabilidad acumulada es igual o mayor que 0.5? Es decir, cuál es el k para el que se cumple la siguiente desigualdad:

$$P(X \leq k) \geq 0.5$$

Puesto que las probabilidades acumuladas se calculan, como acabamos de ver, con `pbinom`, podríamos calcular los valores $P(X \leq k)$ para k de 0 a 10, y buscar, en los resultados, el primero que iguale o supere 0.5. Al contar hay que tener cuidado, y restar 1 porque el primer valor del vector corresponde a $X = 0$ (pero R numera los elementos empezando siempre desde 1).

```
(pAcumuladas = pbinom(0:10, size= 10, prob=0.3))  
  
## [1] 0.02825 0.14931 0.38278 0.64961 0.84973 0.95265 0.98941 0.99841  
## [9] 0.99986 0.99999 1.00000  
  
min(which(pAcumuladas >= 0.5)) - 1  
  
## [1] 3
```

Si te cuesta entender el segundo comando, analízalo paso a paso, ejecutando primero la parte `pAcumuladas >= 0.5`, luego `which(pAcumuladas >= 0.5)`. En general, esa es una buena estrategia si te cuesta entender como está funcionando una expresión complicada en R.

En cualquier caso, el método resulta demasiado complicado como para que lo convirtamos en un recurso habitual. Afortunadamente, ese es precisamente el trabajo que `qbinom` hace por nosotros. Para obtener el resultado anterior, haríamos simplemente:

```
qbinom(0.5, size= 10, prob=0.3)  
  
## [1] 3
```

¿Y si queremos lo contrario? Por ejemplo, vamos a pensar ahora en la binomial $B(7, 1/4)$ (que ya hemos usado antes), y busquemos el *mayor* valor k para el que se cumple:

$$P(X \geq k) \geq 0.75$$

Ejercicio 4.

Haz esta cuenta “a mano”, usando `pbinom` (en particular, sin usar `qbinom`). Solución en la página 59. \square

Si lo deseas, puedes forzar a R a trabajar con la cola derecha. Es decir, puedes obligar a R a utilizar probabilidades de la forma $P(X > k)$ tanto en `pbinom` como en `qbinom`, con la opción `lower.tail=FALSE` (recuerda que, en RStudio, el tabulador es tu amigo).

Ejercicio 5.

Repite la cuenta del Ejercicio 4 usando esa opción. Solución en la página 60. □

Personalmente, te lo desaconsejo. Lo que se puede ganar así, no compensa a mi juicio, lo que se pierde en sencillez conceptual, una vez que te acostumbras a trabajar siempre usando el valor por defecto.

1.5. La función `rbinom`

Cuando conocimos la función `sample` vimos que, usando la opción `probs`, al elegir aleatoriamente en un conjunto de números, podíamos asignar distintas probabilidades para cada uno de los elementos. La función `rbinom` (con `r` de `random`) fabrica números aleatorios pero que, de nuevo, no son equiprobables. Los números que fabrica `rbinom` responden a la distribución de probabilidad binomial que nosotros queramos. Así, por ejemplo, al escribir (no se muestra el resultado):

```
set.seed(2014)
rbinom(n=200, size= 10, prob=0.2)
```

R fabrica un vector de 100 números aleatorios, cuyas probabilidades se ajustan a una distribución binomial $B(10, 0.2)$. En particular, eso significa que sólo podemos obtener resultados de 0 a 10. Y además, como la probabilidad de éxito $p = 0.2$ es baja, es muy poco probable obtener resultados altos, como 9 o 10. Vamos a explorar esto con más detalle.

Ejercicio 6.

1. ¿Cómo de improbable es observar un valor de X mayor o igual a 9 si X es una binomial $B(10, 0.2)$? Calcula $P(X \geq 9)$ para $X \sim B(10, 0.2)$
2. Ya que estamos, calcula $P(X = k)$ para todos los valores de k y representa gráficamente (usa `barplot`) cómo se distribuye la probabilidad.
3. Usando el código que aparece más arriba, fabrica un vector con 200 números aleatorios que sigan la distribución $B(10, 0.2)$. Calcula la media y desviación típica (poblacional y muestral) de esos valores. Compáralos con los valores teóricos.
4. Construye la tabla de frecuencias de los números que has construido, y representa gráficamente (usando `barplot`) esa tabla.
5. Para ver en acción a `rbinom`, ejecuta el siguiente fragmento de código, que te mostrará, a la izquierda, la gráfica de columnas de los valores producidos por `rbinom` y, a la derecha, los valores teóricos de las probabilidades que corresponden a la binomial $B(n, p)$, con los mismos parámetros n y p . Inicialmente, fabricamos 20 valores aleatorios con `rbinom`. Ejecuta el programa varias veces con `numeroDatos = 20`. Después, cambia 20 por 100, y haz lo mismo. Luego, cambia por 1000, 10000 y 1000000 (un millón), y repite en cada caso lo mismo. ¿Qué observas en los gráficos a medida que aumenta `numeroDatos`?

```
numeroDatos = 20
n = 10
p = 0.2
aleatoriosBinomiales = rbinom(numeroDatos, size=n, prob=p)

par(mfrow=c(1,2))
barplot(table(aleatoriosBinomiales)/numeroDatos)
barplot(dbinom(0:n, size=n, prob=p))
par(mfrow=c(1,1))
```

Los comandos `par` son los encargados de colocar las dos figuras juntas.

Solución en la página 60. □

1.6. Media y varianza con R

Esta breve sección tiene por objeto aclarar que no existen funciones específicas para calcular la media y varianza de una distribución binomial en R... porque no se necesitan, claro. Las fórmulas

$$\mu = n \cdot p, \quad \sigma = \sqrt{n \cdot p \cdot q}$$

son tan sencillas que basta con calcular esos valores directamente cuando se necesiten. Vamos a aprovechar, eso sí, para comprobar numéricamente que esas fórmulas coinciden con lo que se obtiene aplicando las definiciones.

Ejercicio 7.

1. Usa R para comprobar que la suma del Ejemplo 5.1.5 (pág. 136) del libro produce los resultados que aparecen en el Ejemplo 5.1.6 (pág. 137). No olvides comprobar también la varianza y desviación típica. Y ten en cuenta que los valores de densidad de la binomial se obtienen con `dbinom` (no es necesario que calcules los números combinatorios). Usa también el panel de vista simbólica de GeoGebra (o Wiris) para comprobar este resultado.
2. El apartado anterior se ha centrado en el cálculo de los valores teóricos de la media μ y varianza σ de una variable binomial $X \sim B(3, 1/5)$. Pero también podemos usar una simulación para comprobar empíricamente ese valor de μ . Y además, y esa es precisamente la novedad que queremos subrayar ahora, la simulación resulta muy fácil si usamos `rbinom`. Compruébalo: usa `rbinom` para fabricar una muestra aleatoria de valores de una variable binomial $X \sim B(3, 1/5)$. Asegúrate de que el tamaño de la muestra es bastante grande (miles, por ejemplo) y calcula la media de los valores de la muestra. Repite este experimento varias veces. ¿Qué observas? Prueba a cambiar el tamaño de la muestra, haciéndola más pequeña. ¿A partir de qué tamaño empiezas a notar diferencias?

Solución en la página 63. □

En el anterior ejercicio hemos evitado a propósito una pregunta que también surge de manera natural: ¿qué sucede con la varianza de los valores de la muestra? Preferimos esperar a que la teoría del capítulo avance un poco antes de volver sobre esta pregunta.

1.7. La binomial con Calc, Wiris y Wolfram Alpha

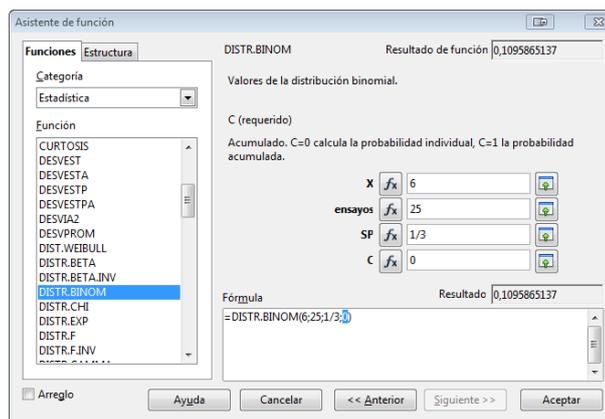
Vamos a presentar muy brevemente las posibilidades de estos programas para trabajar con la binomial. En comparación con lo que ofrece R son, en todos los casos, bastante limitadas.

Calc

Calc ofrece apenas la función `DISTR.BINOM`. Esta función aúna la funcionalidad de `pbinom` y `dbinom`. Es decir, que permite calcular la densidad de probabilidad, pero también la probabilidad acumulada (la función de distribución). En la siguiente figura ilustramos como calcular $P(X = k)$ en una binomial $B(n, p)$, cuando $n = 25$, $p = 1/3$ y $k = 6$. La sintaxis es, como se ve:

`DISTR.BINOM(6;25;1/3;0)`

El último argumento (que Calc llama `C`) sirve para decidir si calculamos los valores de densidad (como en `dbinom`) cuando vale 0, o si calculamos la función de distribución (como en `pbinom`) cuando vale 1



Ejercicio 8.

Repita en Calc los cálculos del Ejercicio 2. O, al menos, repite algunos de ellos y piensa como harías el resto. □

Wiris

Wiris no ofrece herramientas especiales para trabajar con la distribución binomial. En el fichero adjunto:

Tut05-BinomialConWiris.html

hemos incluido una sesión de trabajo con Wiris, en la que se calculan valores de probabilidad de una binomial, y que puedes modificar para cubrir tus necesidades. Como en otros casos, una ventaja de trabajar con Wiris es su capacidad de expresar simbólicamente (como fracciones, sin redondeos) algunos resultados.

Ejercicio 9.

Calcular, usando Geogebra o Wiris, la probabilidad $P(300 < X < 600)$ si $X \sim B(1000, 1/3)$. Es el Ejemplo 5.3.2 del libro (pág. 147), en el que puedes ver la monstruosa solución exacta. Solución en la página 64. □

Wolfram Alpha

Wolfram Alpha es muy potente, pero el mayor problema suele ser el de encontrar la forma correcta de formular la pregunta. Por ejemplo, para calcular $P(X \leq 6)$ cuando $X \sim B(25, \frac{1}{3})$ puedes usar:

Prob x <= 6 if x is binomial with n = 25 and p = 1/3

devuelve el resultado correcto, pero si sustituyes <= por = (para calcular la densidad), Wolfram Alpha no lo entiende... y una forma correcta es esta (mucho menos intuitiva):

PDF[BinomialDistribution[25, 1/3], 6]

donde PDF corresponde a Probability Density Function. No dejes, en cualquier caso, de consultar las páginas de ejemplos de Wolfram Alpha (y una búsqueda astuta en internet a menudo muestra cómo formular la pregunta que nos interesa).

Puedes consultar también esta página web (*Binomial Distribution Calculator*, en *Wolfram Alpha Widgets*),

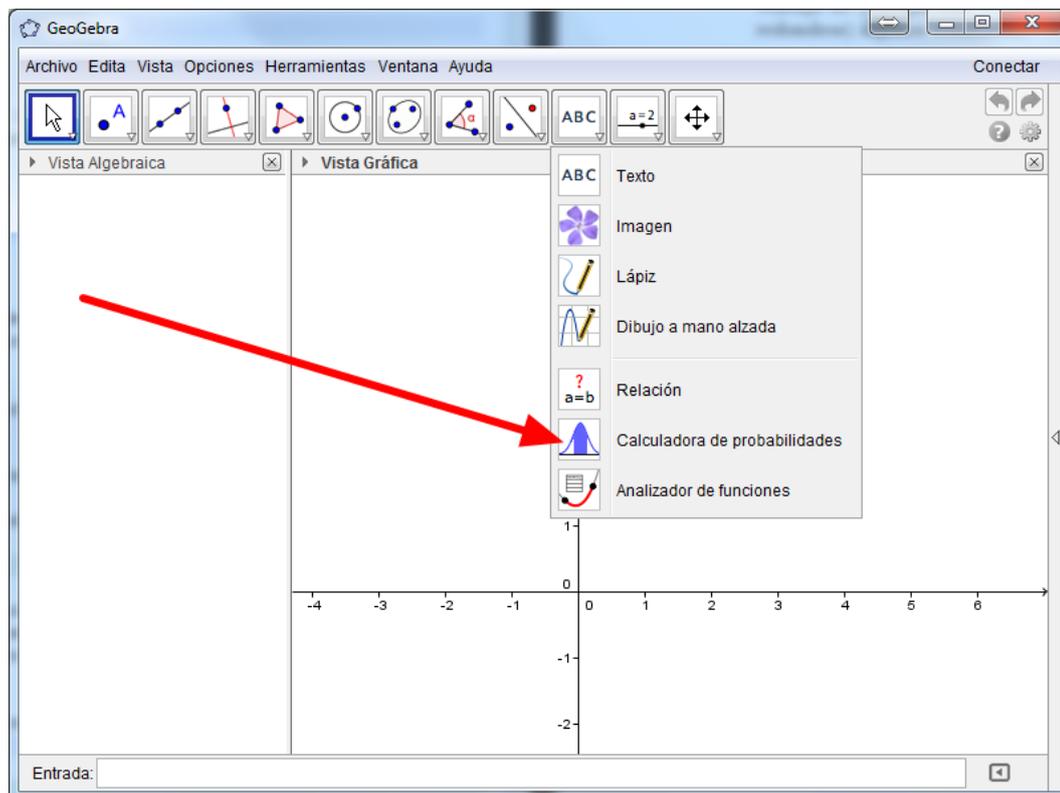
<http://www.wolframalpha.com/widgets/view.jsp?id=78baf4f3a070cc5b9b226664d2ce80ec>

que permite usar Wolfram Alpha para cálculos con la binomial, mediante una interfaz bastante más sencilla.

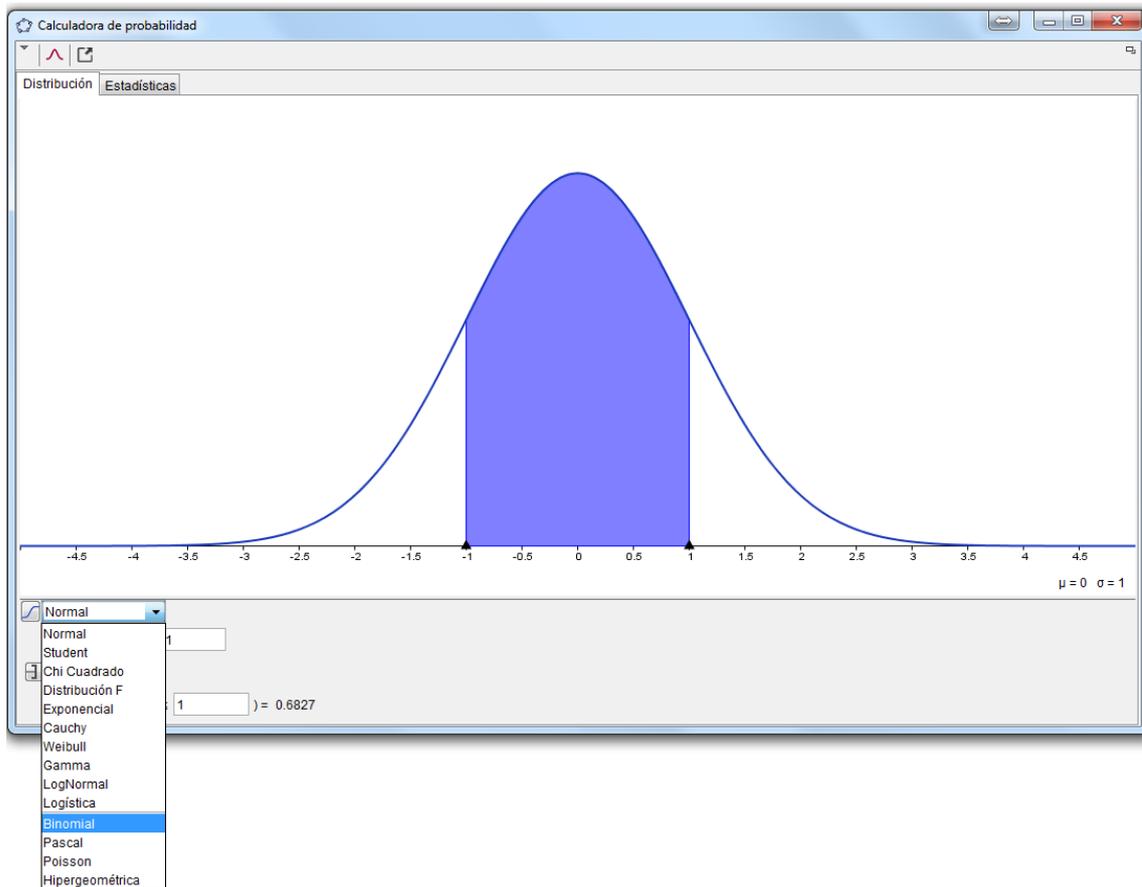
2. Zoo binomial

Vamos a recurrir a GeoGebra para explorar las distintas distribuciones binomiales que se obtienen al variar los valores de n y p . GeoGebra tiene la ventaja de que nos permite hacer esa exploración dinámicamente, de una forma muy sencilla. Además, como veremos, lo vamos a poder utilizar en más casos a lo largo del curso.

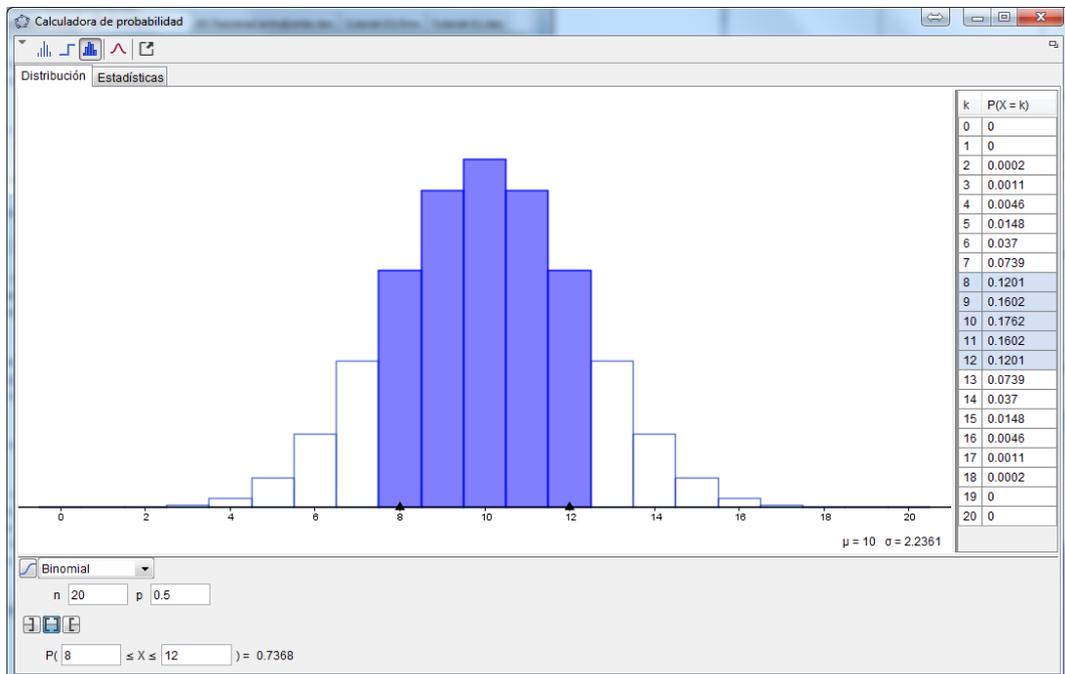
Empezamos abriendo la ventana de GeoGebra (tiene que ser una versión reciente; yo he usado la versión 4.4.39 para este tutorial), y en el menú desplegable de herramientas que se indica en la siguiente figura, seleccionamos la herramienta llamada *Calculadora de Probabilidades*.



Al hacer clic sobre esa herramienta se abrirá una nueva ventana. Te recomiendo maximizarla para aprovechar al máximo el espacio de pantalla. Inicialmente, la ventana muestra una curva en forma de campana, que corresponde a la distribución normal (sobre la que volveremos en breve, tanto en el libro como en este tutorial). Como se ve en la siguiente figura, haz clic sobre el menú desplegable de la parte inferior, que nos permite elegir el tipo de distribución con la que vamos a trabajar, y selecciona la Binomial.



La ventana cambiará de aspecto, para aparecer como en esta figura:



Esta ventana nos permite usar GeoGebra para hacer cálculos como los del Ejercicio 2 (pág. 4 de este tutorial), acompañándolos con la interpretación gráfica de lo que estamos calculando. Para usar la herramienta tenemos que introducir los valores de n y p , junto con los extremos adecuados del intervalo, en los correspondientes campos de esta ventana. Además, usamos los iconos



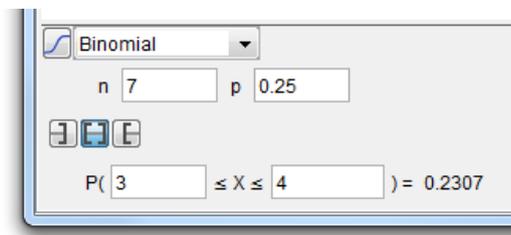
para seleccionar la forma del intervalo que nos interesa, y que se corresponden, respectivamente, con preguntas de uno de estos tipos:

$$P(X \leq a), \quad P(a \leq X \leq b), \quad P(X \geq b)$$

En el caso de la parte 7 del Ejercicio 2 (calcular $P(2 \leq X < 4)$ en $X \sim B(7, 1/4)$), por ejemplo, sería así:



Ampliación de los campos en los que introducimos los valores:



Como ves, en este caso hemos obtenido la respuesta que esperábamos, 0.2307. La ventana, además, ilustra mediante un diagrama de columnas los valores de probabilidad que estamos calculando, mientras que, a la derecha y en vertical, puedes ver la tabla de densidad de la variable X . Si te fijas, verás que GeoGebra ha incluido en la figura los valores de μ y σ . Ya iremos descubriendo algunos elementos adicionales de esta ventana.

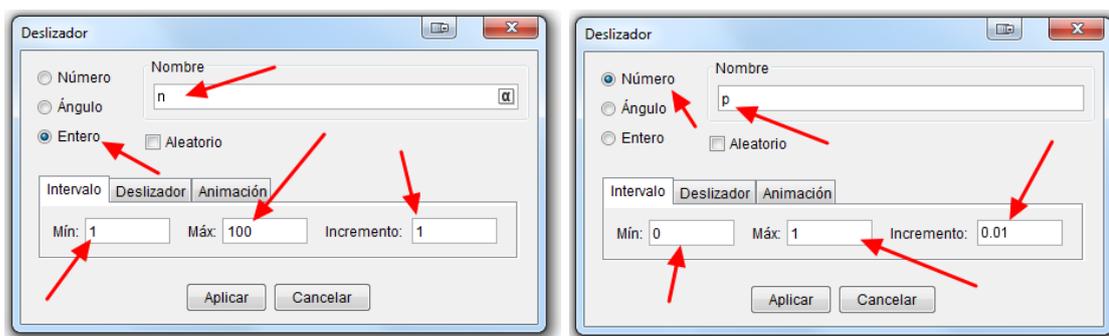
Ejercicio 10.

De nuevo, usa esta herramienta de GeoGebra para reproducir todos los resultados del Ejercicio 2. □

Aunque, en principio, la *Calculadora de Probabilidades* de GeoGebra nos permite examinar el efecto de utilizar distintos valores de n y p , lo podemos hacer mejor volviendo a la interfaz básica del programa. Cierra, por tanto, la *Calculadora* y en la ventana principal de GeoGebra selecciona la

herramienta *deslizador*, haciendo clic en el icono . Ahora usa esa herramienta, haciendo

clic en dos lugares cualesquiera de la vista gráfica, para crear dos deslizadores para n y p . Para crearlos hemos usado las opciones que aparecen indicadas por flechas rojas en estas figuras:



También puedes crear los deslizadores desde la *Línea de entrada* de GeoGebra (que está en la parte inferior de la ventana principal del programa), ejecutando consecutivamente los comandos:

$$n = \text{Deslizador}[1,100,1,0,200]$$

y

$$p = \text{Deslizador}[0,1,0.01,0,200]$$

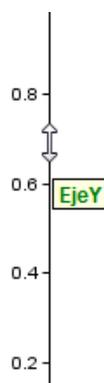
Una vez creados estos deslizadores, usamos de nuevo la *Línea de entrada* y tecleamos (o usamos copia-pegar)

$$\text{DistribuciónBinomial}[n,p]$$

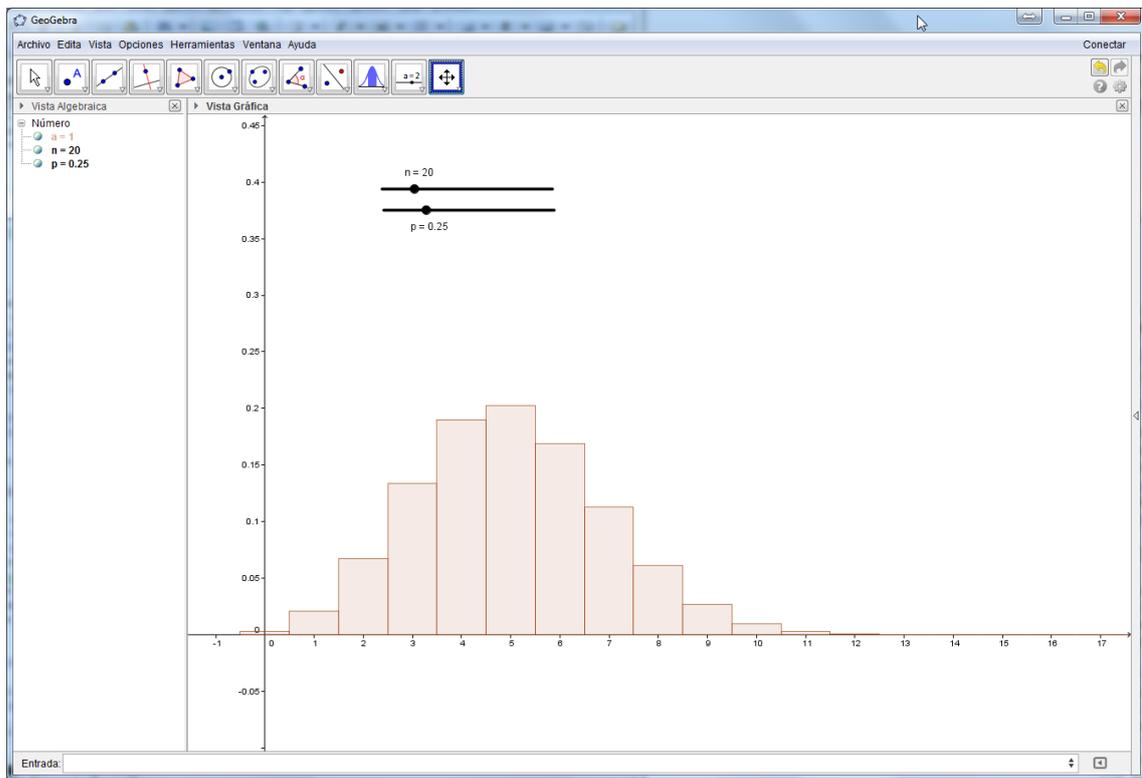
Inicialmente, puesto que tienes $n = 1, p = 1$ la figura no te dirá gran cosa. Para hacerlo interesante, usa el ratón para mover los deslizadores de manera que sea $n = 20$ y $p = 0.25$. Los valores de la probabilidad empiezan a aparecer, pero son poco visibles. El problema, naturalmente, es que al ser valores entre 0 y 1, la escala del eje vertical tiene que cambiar para que podamos apreciarlos. Una posibilidad es ejecutar en la *Línea de entrada* el siguiente comando:

$$\text{RazónEjes}[25,3]$$

Este reajuste de escalas también se puede hacer dinámicamente, usando el ratón. Empieza por hacer clic en el icono  situado en el extremo derecho de la barra de herramientas. Ahora, una vez que hayas seleccionado esa herramienta (que GeoGebra llama *Desplaza Vista Gráfica*), sitúa el ratón sobre el eje vertical. Cuando esté bien colocado el cursor se convertirá en una doble flecha (y puede aparecer un rótulo *EjeY*), como en esta figura:

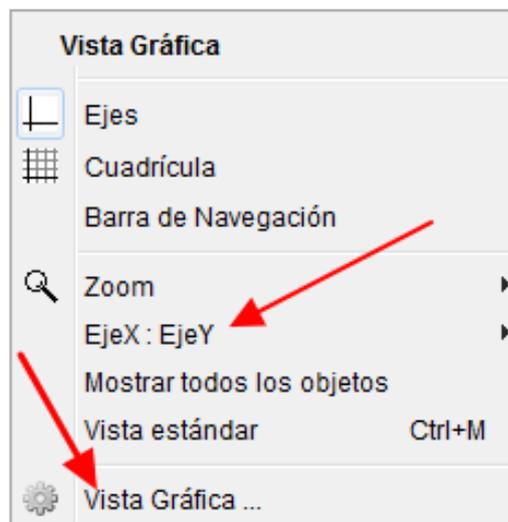


En ese momento haz clic con el botón izquierdo del ratón y, sin soltarlo, “estira” el eje vertical para cambiar su escala. El resultado puede ser algo como lo que se muestra en esta figura.

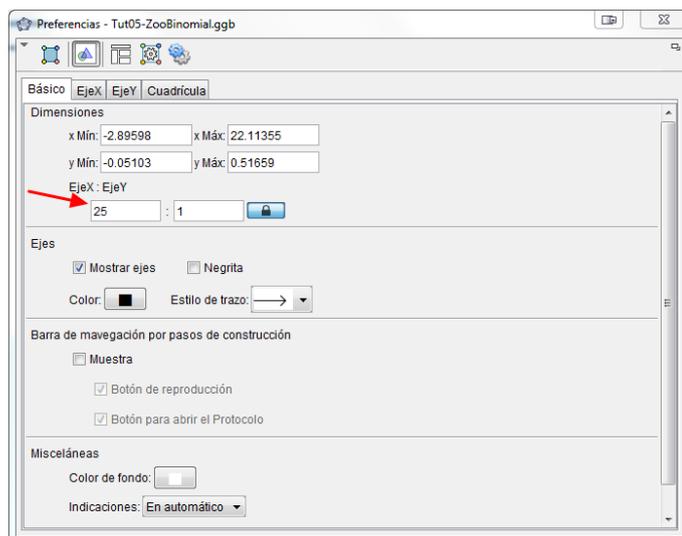


Recuerda, además, que para desplazar la vista gráfica en GeoGebra basta con hacer click (botón izquierdo) en cualquier zona vacía de la *Vista Gráfica* y arrastrar con el ratón, mientras mantienes pulsada la tecla **Ctrl** o la tecla mayúsculas.

Otra posibilidad, una vez que has seleccionado la herramienta *Desplaza Vista Gráfica* es hacer clic con el botón derecho en una zona vacía de la gráfica. Debería aparecer el cuadro de diálogo titulado *Vista Gráfica*, en el que a su vez puedes seleccionar la opción *Vista gráfica*



En ese mismo cuadro de diálogo aparece la opción *EjeX:EjeY*, que te puede servir, pero se limita a ciertos valores prefijados, que no siempre son adecuados a lo que queremos. Tras seleccionar *Vista gráfica*, en la ventana que aparece puedes elegir una relación de escalas entre los ejes similar a 25:1.



y cerrar esta ventana para volver a la vista gráfica. En cualquier caso, por si tienes problemas para ajustar la escala de los ejes, aquí tienes adjunto el fichero

[Tut05-ZooBinomial.ggb](#)

que debería abrirse en GeoGebra mostrando una escala del eje vertical adecuada para que puedas experimentar.

Porque esa es la clave de todo este trabajo de preparación. El objetivo es que juegues con los deslizadores n y p y observes la forma de la distribución de probabilidad. En la Sección 5.1.3 (pág. 137) del libro se describen, a grandes rasgos, tres posibles situaciones que pueden darse en las distribuciones binomiales. Trata de jugar con los valores de n y p hasta asegurarte de que has pasado por esas tres situaciones (para valores de n mayores que 30 es posible que quieras hacer zoom para mejorar la visualización; la rueda del ratón te servirá para esto). La intuición que adquieras sobre el comportamiento de la binomial nos será muy útil para el trabajo de este y los próximos capítulos.

Ejercicio 11.

1. Reproduce, usando los deslizadores n y p las distribuciones binomiales que aparecen en la Figura 5.1 del libro (pág. 139). Si quieres reproducir la Figura 5.2 (o alguna otra binomial fuera del rango de los deslizadores), lo más fácil es que abras una nueva ventana de GeoGebra y utilices directamente el comando `DistribuciónBinomial[10000,0.0001]` (en la Línea de Entrada), además de ajustar la escala de los ejes como hemos descrito.
2. Usa la Calculadora de Probabilidades de GeoGebra para reproducir el resultado que ilustra la Figura 5.6 del libro (pág. 144).

□

3. Variables aleatorias continuas. Funciones e integración con GeoGebra, Wiris y Wolfram Alpha.

En este apartado vamos a aprender a usar algunas de las herramientas que ya conocemos para trabajar con las funciones que usaremos en el estudio de las distribuciones continuas. Para empezar, vamos a practicar el dibujo de gráficas.

3.1. Gráficas de funciones.

Para aprovechar la inercia de la Sección anterior, vamos a empezar usando GeoGebra para la representación de funciones de la forma $y = f(x)$. Por ejemplo, vamos a dibujar la función que aparece en el Ejemplo 5.4.1 del libro (pág. 150).

$$y = \frac{1}{\pi(1+x^2)}$$

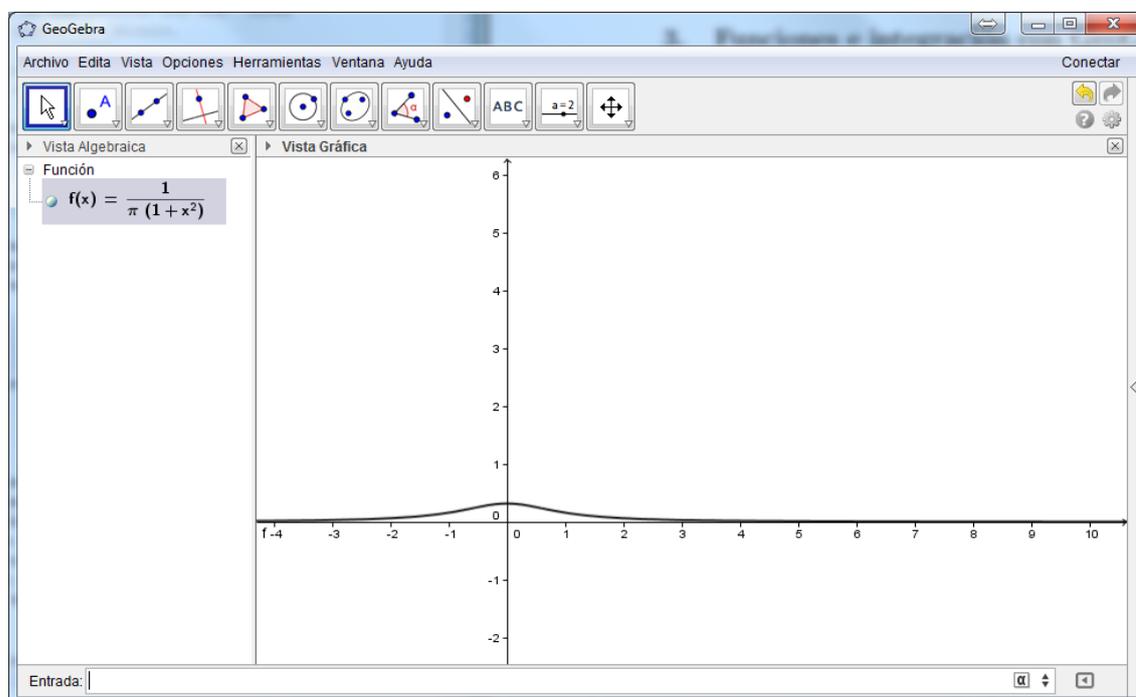
Abre una ventana nueva de GeoGebra (cierra y vuelve a abrir el programa si es necesario) y teclea o copia, en la *Línea de Entrada*, este comando:

$$f(x) = 1 / (\text{Pi} * (1 + x^2))$$

Como ves, es una sintaxis muy parecida a la de R, o Wolfram Alpha (o muchos otros programas de Matemáticas). Hay que usar juiciosamente los paréntesis pero, aparte de eso, la única peculiaridad es que GeoGebra reconoce el símbolo Pi (y también pi en minúsculas) para indicar la constante π . Por ejemplo, en R, sólo funciona pi en minúsculas. Wolfram Alpha normalmente entiende cualquiera de las dos. Ese tipo de detalles cambian de un programa a otro, pero la mayor parte de la sintaxis se mantiene. Y eso facilita muchas veces el mover información de un programa a otro. Uno de los mensajes que queremos transmitir en estos tutoriales es que las herramientas basadas en texto son, en general, más *productivas*, una vez se superan las dificultades iniciales.

Pero, en cualquier caso, cuando uses herramientas basadas en texto para trabajar con funciones, recuerda que estás haciendo una traducción desde la notación matemática habitual a esa notación basada en texto. Es, por tanto, muy importante que te asegures de que la traducción ha sido correcta, y de que estás de hecho trabajando con la función que te interesa. En GeoGebra, por ejemplo, la *Vista Algebraica* te mostrará la función que has introducido en una notación más parecida a la notación matemática habitual.

Cuando hayas ejecutado ese comando GeoGebra, la *Vista Gráfica* tendrá un aspecto similar a este:



Es un buen momento para que practiques con esta figura, ajustando escalas, posición, etc. Experimenta un poco con la gráfica. En este curso nos vamos a limitar a cubrir el conocimiento mínimo necesario sobre funciones. Pero si necesitas más, la documentación de GeoGebra es muy amplia, y la información disponible en la red sobre ese programa es ingente.

Pasando a Wolfram Alpha, prueba a copiar exactamente el mismo comando en el campo de entrada del programa. El comienzo del resultado al ejecutarlo es este:

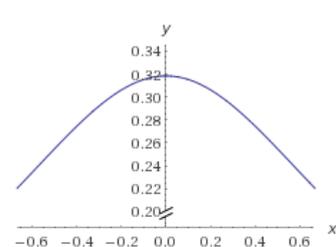
$f(x) = 1 / (\pi * (1 + x^2))$
☆ ☰

☰ 📷 📄 🔄
☰ Examples ➤ Random

Input:

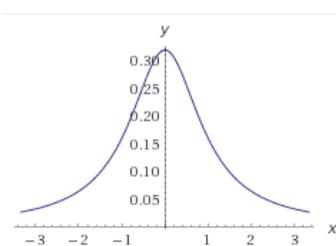
$$f(x) = \frac{1}{\pi(1+x^2)}$$

Plots:



(x from -0.6 to 0.6)

[Enable interactivity](#)



(x from -3 to 3)

[Enable interactivity](#)

Alternate forms:

Además de un par de gráficas a distintas escalas (y de la función en notación matemática, que siempre conviene chequear para ver que el programa nos ha entendido), Wolfram Alpha nos proporciona, como de costumbre, mucha información adicional sobre esta función. Puedes explorarla, para hacerte una idea de lo que obtenemos (por ejemplo, una primitiva).

Ejercicio 12.

Dibuja la gráfica de las siguientes funciones, usando tanto GeoGebra como Wolfram Alpha:

1. La distribución normal de la Ecuación 5.8 del libro (pág. 143), para $\mu = 0$, $\sigma = 1$.
2. Una parábola muy sencilla, $f(x) = x^2$. Después, dibuja (1) $f_1(x) = -x^2$, (2) $f_2(x) = (x-1)^2$, (3) $f_3(x) = (x+1)^2$, (4) $f_4(x) = 2 \cdot x^2$ (5) $f_5(x) = \frac{1}{2}x^2$.

El objetivo del anterior ejercicio, como el de este, es hacerte pensar sobre lo que sucede con la gráfica de una función $y = f(x)$ al hacer un cambio de variable de la forma

$$y = c \cdot f(a \cdot x + b),$$

donde a, b y c son constantes.

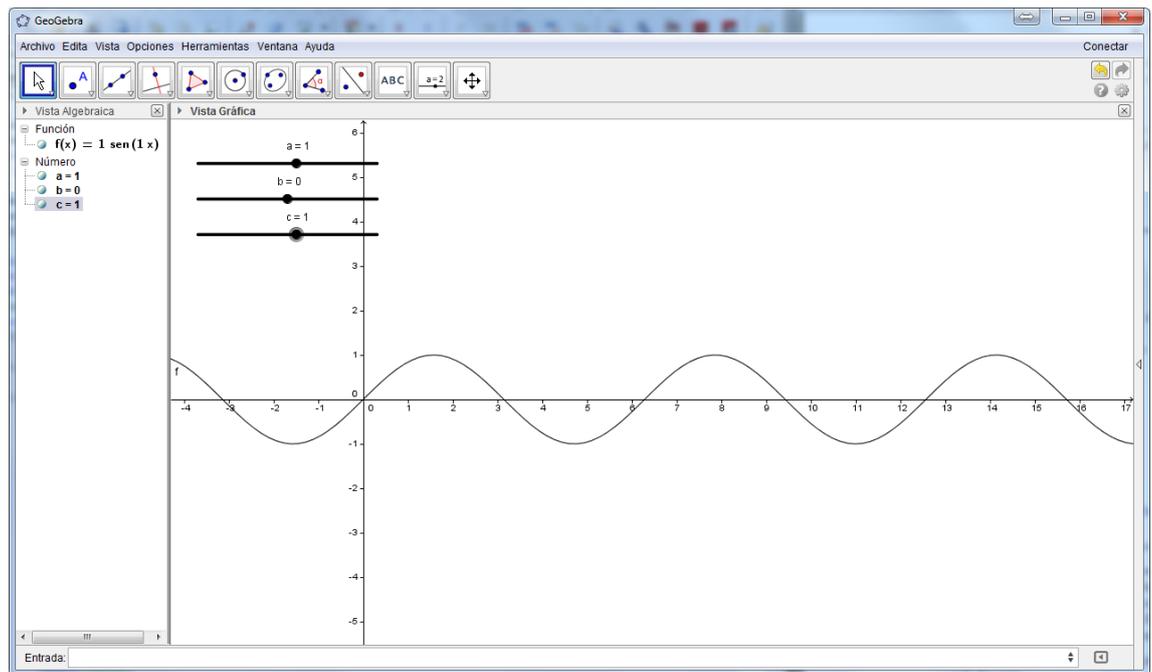
3. Para profundizar en esto vamos a usar los deslizadores de GeoGebra. Abre una nueva ventana de GeoGebra y escribe, uno por uno, estos comandos:

```
a = Deslizador[-10,10,0.1,0,200]
a = 1
b = Deslizador[-10,10,0.1,0,200]
c = Deslizador[-10,10,0.1,0,200]
c = 1
```

Con esto hemos definido tres deslizadores, cuyo recorrido va de -10 a 10 , en incrementos de 0.1 , y hemos fijado el valor inicial de a y c a 1 (el valor inicial de b es 0). No te preocupes por las dos últimas opciones del comando **Deslizador**, que no vamos a usar. Naturalmente, también puedes usar el icono  de la Barra de Herramientas. Ahora ejecuta el comando:

$$f(x) = c * \text{sen}(a * x + b)$$

En ese momento, el aspecto de la ventana gráfica debe ser similar a este:



Prueba a manipular los deslizadores para ver el efecto que los cambios en a , b y c tienen sobre la gráfica. Los tres efectos que se observan son fáciles de traducir en palabras. Inténtalo.

4. Hay un cuarto efecto que puedes analizar, para el que vamos a añadir un deslizador adicional, llamado d , que corresponde a esta fórmula:

$$y = c \cdot f(a \cdot x + b) + d.$$

Usa el ejemplo de la función seno para añadir un deslizador más, que represente el valor de d , y estudiar su efecto en la gráfica. Tendrás que redefinir la función f en GeoGebra, usando de nuevo la Línea de Entrada.

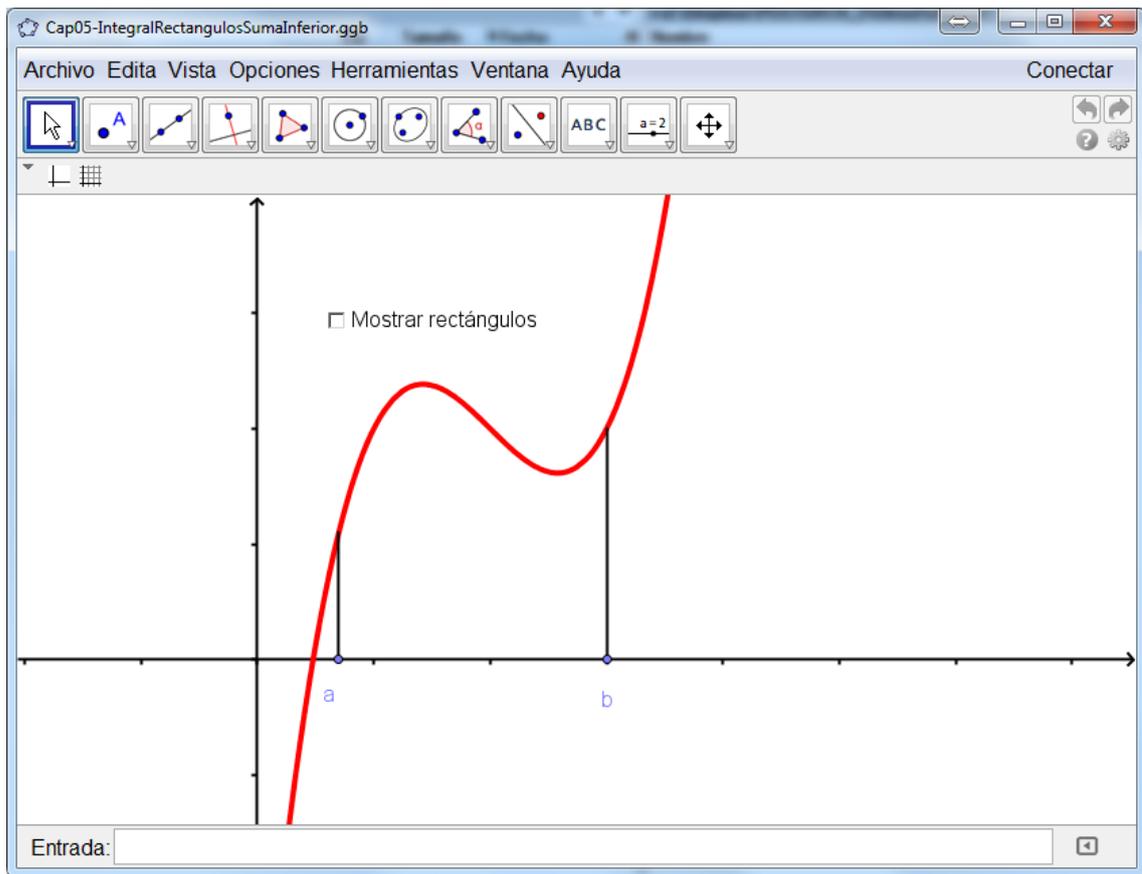
□

3.2. Integrales y sumas de áreas de rectángulos.

La Figura 5.7 del libro (pág. 146) trata de ilustrar que la idea de integral aparece al pensar en la aproximación del área bajo una curva mediante la suma del área de muchos rectángulos inscritos bajo esa gráfica. Esa figura se ha obtenido con el fichero GeoGebra

[Tut05-IntegralRectangulosSumaInferior.ggb](#)

que, además, nos va a permitir explorar de forma dinámica lo que ocurre con esas aproximaciones cuando n aumenta. Al abrir el fichero te encontrarás en la *Vista Gráfica* una situación parecida a esta:



Haz clic sobre la casilla rotulada *Mostrar rectángulos*, para hacer aparecer unos rectángulos inscritos bajo la gráfica de f (inicialmente 2), y un deslizador que controla el número de rectángulos. En la Figura 5.7 ese deslizador se ha usado para obtener 15 rectángulos. Experimenta con ese deslizador, para que veas lo que sucede con el área de los rectángulos a medida que aumenta su número. En la siguiente figura hemos hecho zoom y desplazado la gráfica para que veas con un poco más de detalle lo que sucede cuando usamos 100 rectángulos:



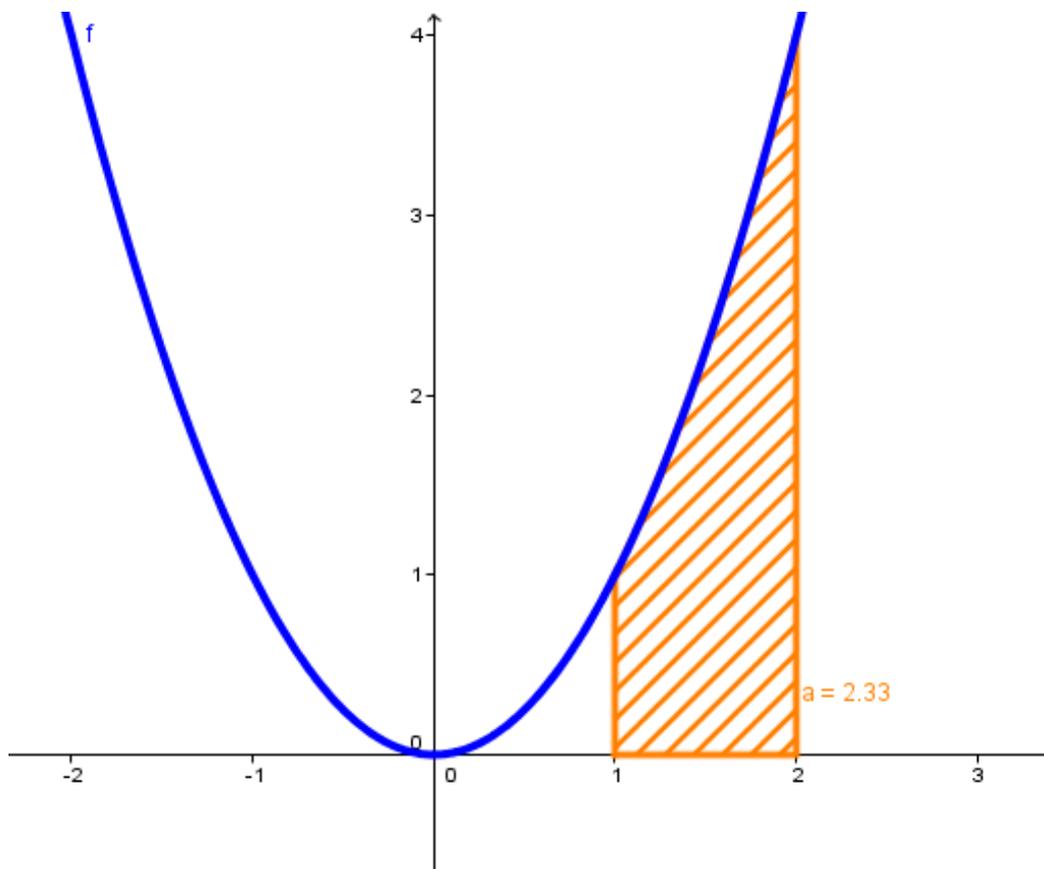
La diferencia entre la suma del área de los rectángulos y el área bajo la gráfica es, como ves, muy pequeña. Por si deseas usarlo para otros experimentos, el comando de GeoGebra que hemos usado para dibujar esos rectángulos es *SumaInferior*. Puedes buscarlo en la ayuda de la página www.geogebra.org.

3.3. Cálculo de áreas y primitivas con GeoGebra, Wolfram Alpha y Wiris.

En el Capítulo 5 del libro se discute que, para trabajar con variables aleatorias continuas, es necesario el cálculo de áreas mediante integrales, y que una de las herramientas básicas para hacer esto es el cálculo de primitivas. Tradicionalmente, el cálculo manual de primitivas ha venido siendo una de las partes más impopulares del estudio de las Matemáticas. Afortunadamente, la aparición de los sistemas de cálculo computacional (tanto simbólicos como numéricos) hacen que la parte más complicada de esos cálculos podamos dejarla, casi siempre, en manos del ordenador. Desde luego, así va a ser en este curso. En esta sección vamos a ver como usar GeoGebra, Wiris y Wolfram Alpha para resolver algunos de los problemas que hemos ido encontrando en ese capítulo.

3.3.1. Un primer ejemplo usando integrales.

Vamos a empezar con un ejemplo muy sencillo, que no está directamente relacionado con el cálculo de probabilidades, pero que nos servirá para hacernos una idea de lo que vamos a hacer. Consideremos la parábola $y = x^2$, y queremos calcular el área que queda debajo de esta parábola y por encima del eje x , entre los puntos de coordenadas $x = 1$ y $x = 2$. El problema se ilustra en esta figura, obtenida con GeoGebra (puedes ver que, de hecho, GeoGebra nos adelanta la respuesta aproximada, que es 2.333):



Para situarlo en un contexto histórico, podemos decir que este problema es, seguramente, el tipo más complicado de problemas de áreas que los matemáticos griegos de la era clásica consiguieron resolver, y que desde ellos hasta Newton no se produjeron grandes avances en ese problema.

Para nosotros las cosas son extremadamente sencillas. Por ejemplo, usando Wolfram Alpha podemos hacer:

```
integrate x^2 from 1 to 2
```

y obtenemos la respuesta que se muestra en la siguiente figura:

Enter what you want to calculate or know about:

integrate x^2 from 1 to 2



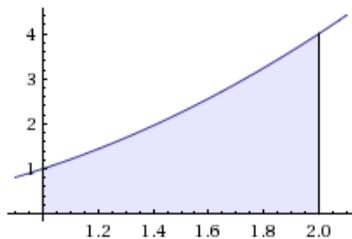
Examples ↻ Random

Definite integral:

More digits

$$\int_1^2 x^2 dx = \frac{7}{3} \approx 2.33333$$

Visual representation of the integral:



Riemann sums:

More cases

left sum	$\frac{14n^2 - 9n + 1}{6n^2} = \frac{7}{3} - \frac{3}{2n} + O\left(\left(\frac{1}{n}\right)^2\right)$
----------	---

Enable interactivity

Indefinite integral:

Approximate form

Step-by-step solution

$$\int x^2 dx = \frac{x^3}{3} + \text{constant}$$

Como ves, el resultado es $\frac{7}{3} \approx 2.3333$, como esperábamos. Además, Wolfram Alpha nos proporciona también una primitiva de $f(x)$, en el bloque titulado *Indefinite Integral*. Concretamente:

$$F(x) = \int x^2 dx = \frac{x^3}{3} + c$$

siendo c una constante. Esta primitiva nos sirve para obtener por otro medio la integral (el área) anterior, usando la Ecuación 5.10 del libro (pág. 151) en lo que se conoce como Regla de Barrow:

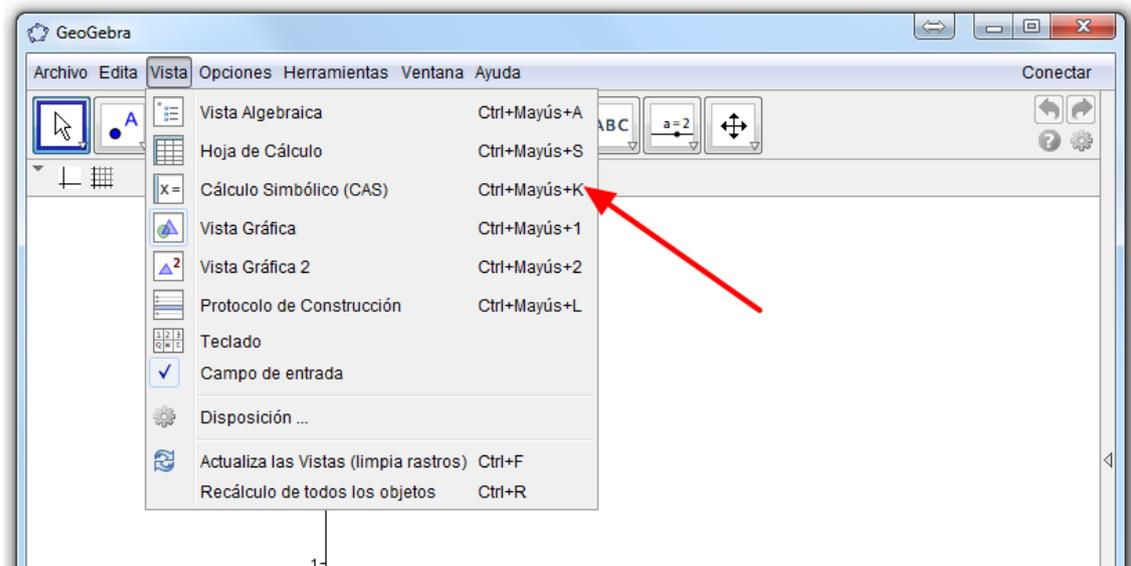
$$\int_1^2 f(x) = F(2) - F(1) = \left(\frac{2^3}{3} + c\right) - \left(\frac{1^3}{3} + c\right) =$$

(como ves la constante c se cancela en este cálculo)

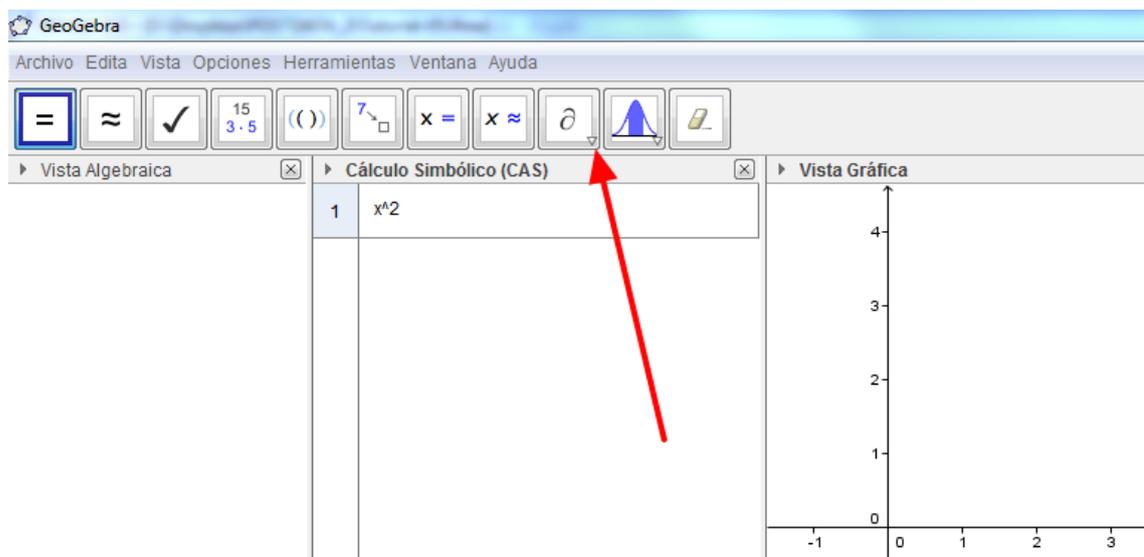
$$= \frac{8}{3} - \frac{1}{3} = \frac{7}{3},$$

como ya nos había dicho Wolfram Alpha.

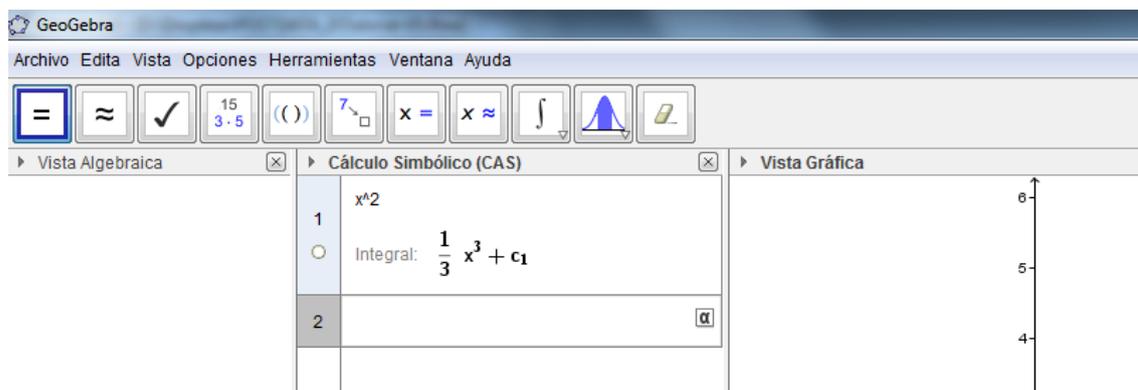
En GeoGebra, para conseguir lo mismo vamos a usar una nueva herramienta, llamada *Cálculo Simbólico (CAS)*. La encontrarás en el menú *Vista*, como indica esta figura:



Al usar esta opción, en la ventana de GeoGebra aparecerá un nuevo panel. Te aconsejo que maximices la pantalla para disponer de espacio suficiente. Sitúate en el campo indicado con un 1, teclea x^2 y **todavía no pulses Entrar**. Tu ventana tendrá este aspecto (sólo se muestra una parte):



Ahora hay que ir con cuidado: haz clic en el pequeño triángulo de la esquina inferior derecha del icono  de la *Barra de Herramientas*. Tiene que ser en el triángulo para desplegar el menú, porque si haces clic directamente en el centro del icono, calcularás la derivada en lugar de la primitiva, y tendríamos que volver atrás. Cuando aparezca el menú desplegable haz clic en la opción *Integral*, del icono . El resultado debería ser similar a este:



Puedes usar copiar y pegar (botón derecho del ratón, recuerda) para trasladar el resultado a otros programas, con las precauciones sintácticas habituales en esos casos. GeoGebra, como ves, es un poco más complicado de usar que Wolfram Alpha para el cálculo de primitivas. A cambio, puedes utilizarlo sin conexión a internet, y nos consta que cada nueva versión ha supuesto una mejora considerable en las posibilidades de uso del programa (en el momento de escribir esto, estamos esperando la versión 5 que, por ejemplo, incluirá gráficos tridimensionales).

En las próximas secciones vamos a usar también Wiris para algunas de estas operaciones, y explicaremos cómo se calculan las primitivas en ese programa.

3.3.2. Una función de densidad de una variable continua.

Empecemos con el Ejemplo 5.4.1 del libro (pág. 150), en el que hemos dejado pendiente la tarea de verificar que la función

$$f(x) = \frac{1}{\pi(1+x^2)}$$

satisface

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

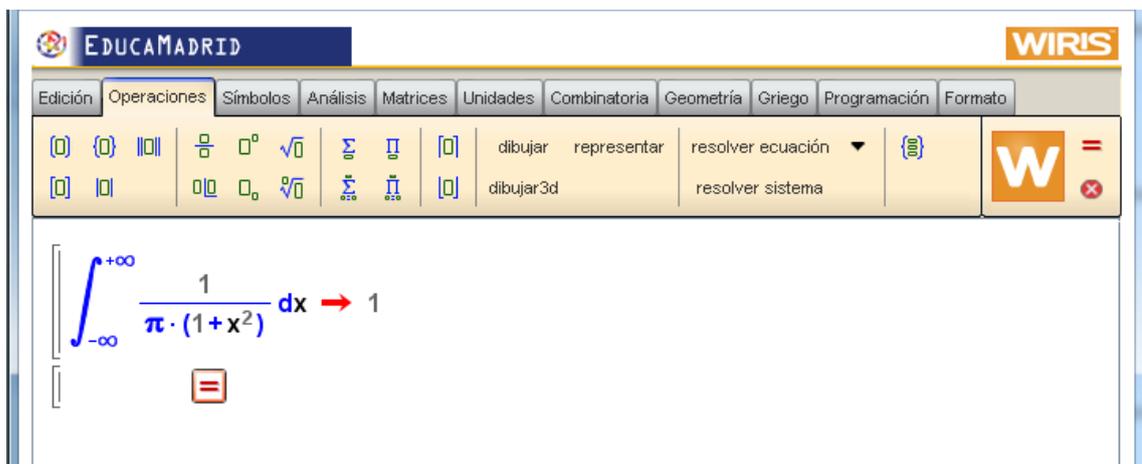
(Ver también el Ejemplo 5.4.3, pág. 153). Para comprobar esto, empezaremos usando Wiris. Ve

a la pestaña **Análisis** y, utilizando el icono  y otros símbolos que ya conoces, escribe

la fórmula como se ve en la siguiente figura (necesitarás π y los símbolos de infinito positivo y negativo, de la pestaña **Símbolos**; no confundas π con la letra del mismo nombre de la pestaña **Griego**).

$$\left[\int_{-\infty}^{+\infty} \frac{1}{\pi \cdot (1+x^2)} dx \right]$$

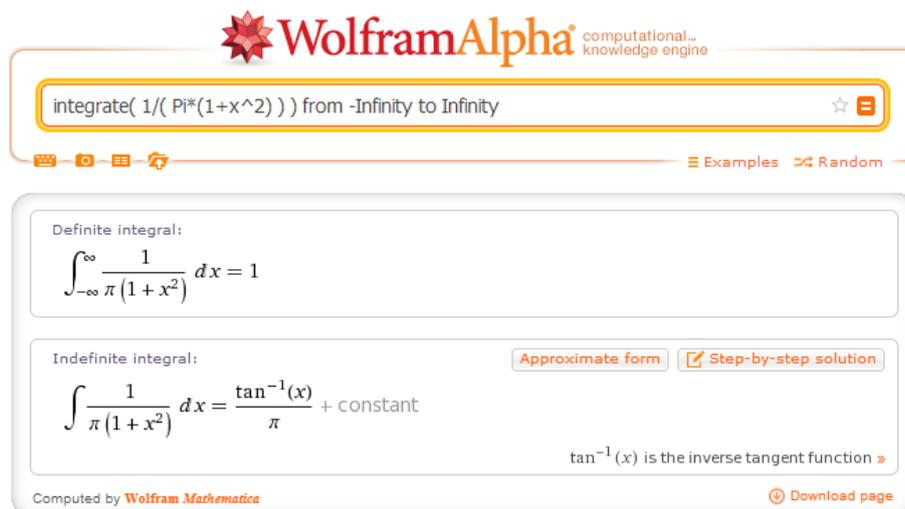
No olvides el símbolo de producto entre π y el paréntesis del denominador, ni la variable x en dx . Después pulsa sobre el símbolo igual, y en unos segundos, Wiris te confirmará que el resultado de la integral es, como esperábamos, 1.



Para hacer lo mismo con Wolfram Alpha, teclea en el campo de entrada,

`integrate(1/(Pi*(1+x^2))) from -Infinity to Infinity`

y, de nuevo, en unos instantes tendrás la confirmación del resultado, como se ve en esta figura.



De hecho, Wolfram Alpha nos da más información de lo que hemos pedido, y nos resuelve la siguiente tarea que dejamos pendiente en el libro: la búsqueda de una primitiva (o integral indefinida) de $f(x)$. Como puede verse en la anterior figura, Wolfram Alpha nos informa de que esa primitiva es

$$\int \frac{1}{\pi \cdot (1 + x^2)} dx = \frac{\tan^{-1}(x)}{\pi} + constant$$

En esta notación (algo confusa) el símbolo \tan^{-1} se refiere al arcotangente (que, en otros programas, se representa con menos ambigüedad mediante **atan**). Para obtener esa primitiva en Wiris vuelve a la pestaña análisis, selecciona esta vez el símbolo (integral sin límites, o indefinida)



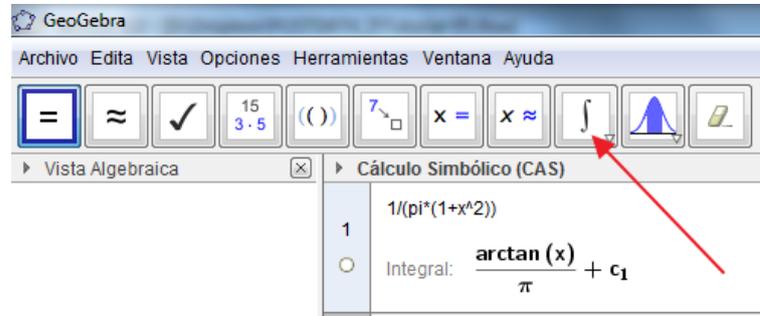
y úsalo (puedes copiar y pegar trozos de la anterior integral, para facilitarte el trabajo), para obtener el mismo resultado expresado de otra manera:

$$\int \frac{1}{\pi \cdot (1 + x^2)} dx \rightarrow \frac{\text{atan}(x)}{\pi}$$

(Wiris toma 0 como valor de la constante que aparece en Wolfram Alpha). Y para terminar el recorrido por los programas que venimos usando, en el panel de cálculo simbólico de GeoGebra podemos comprobar el resultado de la integral con el comando:

$$\text{Integral}[1 / (\text{pi} * (1 + x^2)), -\infty, \infty]$$

mientras que la primitiva se obtiene así:



3.3.3. Probabilidad de un intervalo.

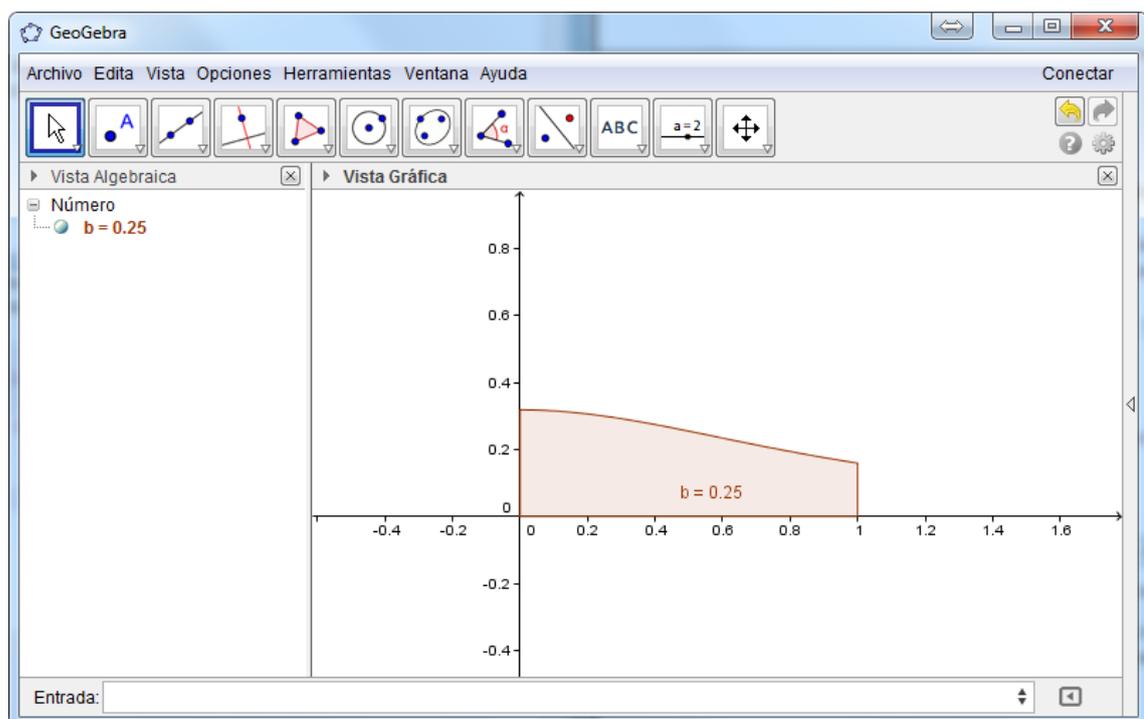
Para completar los resultados de los Ejemplos 5.4.1 y 5.4.2 del libro (pág. 150), todavía tenemos que calcular la probabilidad del intervalo $[0, 1]$. Es decir:

$$P(0 \leq X \leq 1) = \int_0^1 f(x)dx = \int_0^1 \frac{1}{\pi(1+x^2)} dx,$$

Vamos a empezar esta vez por GeoGebra. Escribe y ejecuta en la *Línea de Entrada* el siguiente comando:

$$\text{Integral}[1 / (\text{pi} * (1 + x^2)), 0, 1]$$

El resultado se muestra gráficamente, como ves en la figura:



Aunque también aparece a la izquierda, en la *Vista Algebraica*. Y si lo escribes en el panel de Cálculo Simbólico obtendrás una respuesta simbólica. ¡Pruébalo!

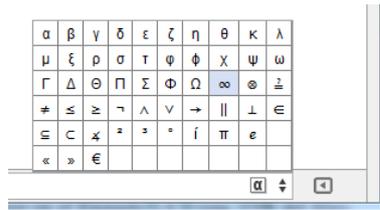
Continuando con esta misma variable aleatoria continuas, en el Ejemplo 5.4.4 del libro (pág. 154) hemos usado la primitiva de esta función para comprobar que la probabilidad

$$P(X \geq 1) = \int_1^{\infty} f(x)dx = \int_0^1 \frac{1}{\pi(1+x^2)} dx,$$

es igual a $\frac{1}{4}$. Vamos a usar GeoGebra para calcular directamente esa integral. La cuenta es muy parecida a la anterior, pero el comando ahora es:

Integral[1 / (pi * (1 + x^2)), 1, ∞]

Si quieres teclear directamente este comando, necesitarás el símbolo de infinito. Para obtenerlo, puedes usar la paleta de símbolos de GeoGebra, que aparece al hacer clic sobre el icono  situado a la derecha de la *Línea de Entrada*.



Ejercicio 13.

1. Repite el cálculo de las integrales

$$\int_0^1 \frac{1}{\pi(1+x^2)} dx,$$

y

$$\int_1^{\infty} \frac{1}{\pi(1+x^2)} dx,$$

con Wolfram Alpha y Wiris.

2. Calcula, usando Wiris, el valor de la integral

$$\int_{\frac{1}{2}}^{\frac{3}{4}} 6 \cdot (x - x^2) dx$$

que aparece en el Ejemplo 5.4.5 (pág. 157) del libro.

3. Calcula, también con Wiris, una primitiva (o integral indefinida) de la función $f(x) = 6 \cdot (x - x^2)$.
4. Comprueba, usando GeoGebra, que $f(x) = 6 \cdot (x - x^2)$, con soporte en el intervalo $0 \leq x \leq 1$ (la función vale 0 fuera de ese intervalo) es una función de densidad. Es decir, comprueba que:

$$\int_0^1 6 \cdot (x - x^2) dx = 1$$

5. Repite el apartado anterior con Wolfram Alpha o Wiris.

Soluciones en la página 64. □

3.4. Funciones definidas a trozos con GeoGebra

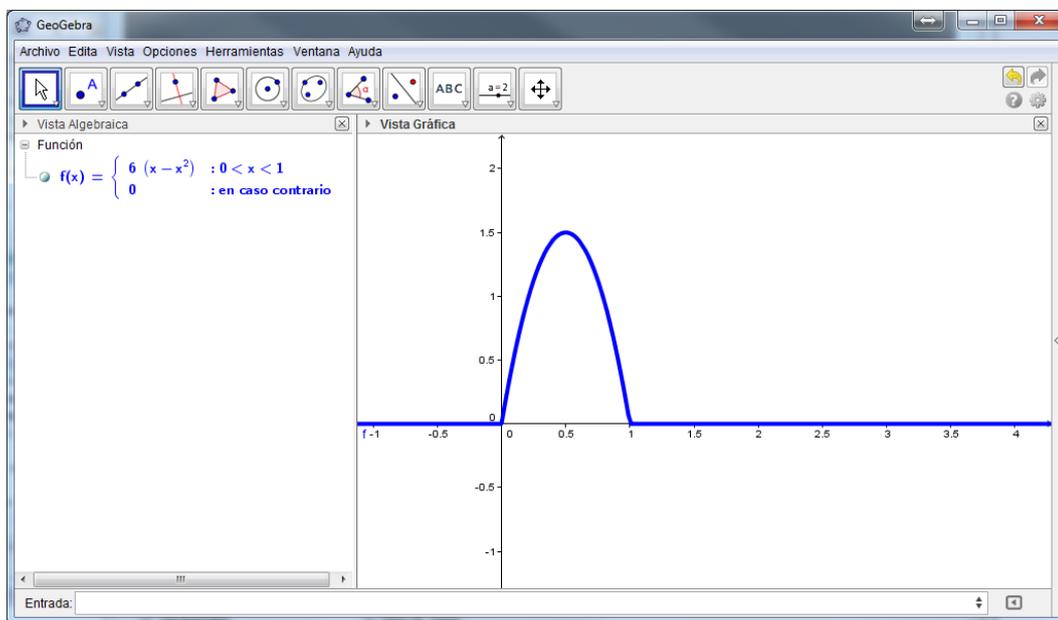
Cuando trabajamos con variables aleatorias continuas con soporte en un intervalo (ver la Sección 5.4.1 del libro, pág. 156), tenemos que usar funciones de densidad definidas a trozos (mediante llaves), como en la función del Ejemplo 5.4.5 de esa sección:

$$f(x) = \begin{cases} 6 \cdot (x - x^2), & \text{para } 0 \leq x \leq 1. \\ 0, & \text{en otro caso.} \end{cases}$$

Esta breve sección sirve sólo para hacer justicia al excelente trabajo que han hecho los programadores de las últimas versiones de GeoGebra, para que el programa sea capaz de trabajar adecuadamente con este tipo de funciones. La clave es la instrucción *Si*, que permite que el resultado de una función dependa del cumplimiento de una cierta condición. Por ejemplo, para definir en GeoGebra la anterior función f basta con teclear, en la *Línea de Entrada* de GeoGebra, el siguiente comando:

$$f(x) = \text{Si}[0 < x < 1, 6*(x - x^2), 0]$$

El resultado se muestra en esta figura:



Hemos coloreado de azul la gráfica para que veas que la función está definida para todos los valores de x , no sólo para los del intervalo $[0, 1]$. Fíjate también en que la descripción de la función que hace GeoGebra en el panel de *Vista Algebraica* (a la izquierda) es impecable. Realmente los programadores han hecho un gran trabajo. Que se traduce, entre otras cosas, en que el cálculo de probabilidades para esta variable se puede realizar de forma muy cómoda. Si, por ejemplo, queremos calcular

$$P\left(\frac{3}{7} < X < \frac{23}{15}\right)$$

podemos traducir directamente los extremos del intervalo a límites de integración, escribiendo:

$$\text{Integral}[f(x), 3/7, 23/15]$$

sin preocuparnos de si esos extremos caen dentro del intervalo $[0, 1]$, o no.

Ejercicio 14.

1. Usa ese comando en GeoGebra para calcular la probabilidad $P\left(\frac{3}{7} < X < \frac{23}{15}\right)$.
2. Calcula la probabilidad $P(-4, 0.7)$ para la misma variable aleatoria X .
3. Este ejercicio es un poco más complicado. Supongamos que ahora Y es una variable aleatoria continua, con soporte en el intervalo $[1, 4]$, dada por esta función de densidad:

$$g(x) = \begin{cases} \frac{2(x-1)}{3} & \text{si } 1 \leq x < 2 \\ \frac{4-x}{3} & \text{si } 2 \leq x < 4 \\ 0 & \text{en cualquier otro caso.} \end{cases}$$

Empieza por comprobar que $g(x)$ es realmente una función de densidad. Una vez hecho esto, calcula la probabilidad:

$$P\left(\frac{1879}{1250} < Y < \frac{2511}{625}\right).$$

3.5. Media y varianza de variables aleatorias continuas.

Si hemos conseguido perderle el miedo a las integrales, las fórmulas para el cálculo de medias y varianzas son muy fáciles de usar. En el Ejemplo 5.4.6 del libro (pág. 162) hemos calculado la media y varianza de una variable aleatoria mediante las integrales:

$$\mu = \int_1^2 x f(x) dx = \int_1^2 x \cdot 2 \cdot (2 - x) dx = \frac{4}{3} \approx 1.33$$

y

$$\sigma^2 = \int_1^2 (x - \mu)^2 f(x) dx = 2 \int_1^2 \left(x - \frac{4}{3}\right)^2 (2 - x) dx = \frac{1}{18} \approx 0.055556$$

respectivamente. Comprobar estos resultados, con cualquiera de los programas que venimos utilizando, es muy sencillo. Por ejemplo, para calcular la media usando GeoGebra basta con ejecutar estos comandos, uno tras otro:

```
f(x) = 2 * (2 - x)
mu = Integral[ x * f(x), 1, 2]
sigma = Integral[ (x - mu)^2 * f(x), 1, 2]
```

Como de costumbre en GeoGebra se muestran los resultados numéricos, no simbólicos. Usa la opción *Redondeo* del menú *Opciones*, si deseas ver más cifras significativas.

Ejercicio 15.

1. Repite estas integrales con *Wiris* y *Wolfram Alpha*.
2. Calcula la media y varianza de la función del Ejemplo 5.4.5 del libro (pág. 157), que ya hemos usado antes en este tutorial, al principio de la Sección 3.4 (pág. 25).
3. Hemos empezado nuestro trabajo sobre variables aleatorias continuas hablando de la función de densidad

$$f(x) = \frac{1}{\pi(1+x^2)},$$

que define la que se conoce como distribución de Cauchy. Y dijimos que era una función simple, pero no engañosamente simple. Para entender en qué sentido este ejemplo no es trivial del todo, trata de calcular su media, con cualquiera de los programas que hemos usado. Comprobarás que la media no existe, porque la integral

$$\int_{-\infty}^{\infty} x \cdot \frac{1}{\pi(1+x^2)} dx$$

es una indeterminación de la forma $\infty - \infty$. En concreto, la integral se puede escribir como una suma:

$$\int_{-\infty}^0 x \cdot \frac{1}{\pi(1+x^2)} dx + \int_0^{\infty} x \cdot \frac{1}{\pi(1+x^2)} dx,$$

en la que la primera integral vale $-\infty$ y la segunda vale ∞ . Usa el ordenador para comprobar esto.

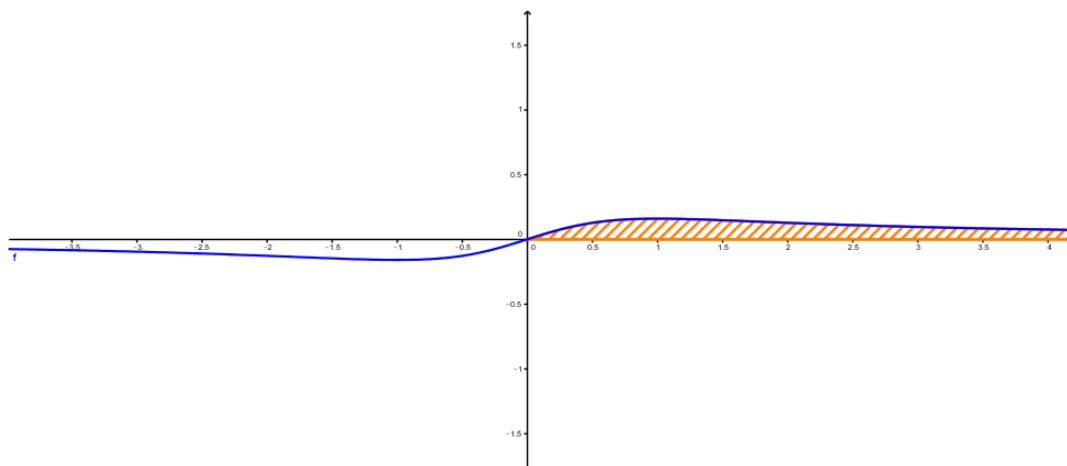
Soluciones en la página 66.

□

Observaciones sobre el último apartado de este ejercicio: El resultado del último apartado plantea algunas dificultades teóricas. No te preocupes demasiado si no lo entiendes del todo la primera vez que lo leas. Es una buena idea usar GeoGebra o Wolfram Alpha para dibujar la gráfica de

$$h(x) = x \cdot \frac{1}{\pi(1+x^2)},$$

como hemos hecho nosotros en la siguiente figura. La simetría de las dos mitades de esa gráfica ayuda a entender porque el resultado de las dos integrales es el mismo, pero con signos opuestos. Y el área de la mitad derecha, la zona sombreada de la figura, es igual a ∞ . Un matemático diría que la integral de 0 a ∞ es divergente.



¿Por qué es divergente esta integral? Es decir, ¿por qué ese área es igual a ∞ ? Si miras la figura verás que la *base* de la figura es el intervalo $(0, +\infty)$, mientras que la altura, dada por la función $h(x)$, se hace más y más pequeña (tiende a 0) a medida que x aumenta. Si piensas en una fórmula del tipo “base \cdot altura”, lo que sucede aquí es que la base tiende a ∞ , mientras la altura tiende a 0. Así que el área es $\infty \cdot 0$, una indeterminación. Las indeterminaciones, en las que dos cantidades exhiben comportamientos enfrentados, se resuelven en Matemáticas estudiando la *velocidad* relativa a la que se mueven esas cantidades. Pero medir esas velocidades y detectar las diferencias entre ellas, que a veces son sutiles, puede ser un trabajo muy complicado. En el Tutorial03 nos encontramos con una situación muy similar al hablar de *series* (sumas de infinitos números). Y allí ya dijimos que la clave era el análisis de la velocidad a la que cambiaban las cantidades involucradas. A menudo, para resolver el problema satisfactoriamente se hace necesario consultar con un experto, un matemático. Una consecuencia importante de esas dificultades teóricas es que los programas de ordenador a veces se ven en un aprieto para calcular este tipo de integrales. En el menos malo de los casos, el programa nos dirá algo así como “no sé calcular esta integral”, y al menos sabremos que tenemos un problema. Pero en ocasiones, las decisiones que hay que tomar en el cálculo de las integrales son tan complicadas que los programas fallan, y producen un resultado erróneo. Veremos un ejemplo en la Sección

¿Qué consecuencias tiene el hecho de que no exista la media? Todavía no podemos contestar con propiedad a esta pregunta, hasta que no hayamos discutido algo sobre distribuciones muestrales en el Capítulo 6 del libro. Pero te podemos adelantar algo: la media se ha diseñado para ser un “*buen representante*” de un conjunto de datos. Si no existe la media, quiere decir que, por alguna razón, no es posible elegir un buen representante para esos datos. Y, a otro nivel, si no existe la media desde luego tampoco existe la varianza.

3.6. La variable de integración.

Más adelante, en la Sección 5.4.3 (pág. 163) del libro, hemos necesitado calcular una primitiva de la función

$$f(x) = \begin{cases} k & \text{si } a < x < b \\ 0 & \text{en otro caso.} \end{cases}$$

donde k es una constante. Nuestro objetivo aquí es encontrar cuál es el valor de k que hace que esta función sea una función de densidad y, para eso, se necesita que su integral en el intervalo $[a, b]$ sea igual a 1.

La dificultad es que, si usamos un programa como GeoGebra, Wiris o Wolfram Alpha para calcular esa primitiva, tenemos que asegurarnos de que el programa entiende que la variable de la función es x y no k . Nosotros sabemos que queremos pensar en x como una variable y k como una constante, pero si no le proporcionamos más información, entonces para el programa x y k son dos símbolos cualesquiera, y puede interpretar ambos como variables.

Wiris.

En Wiris basta, en principio, con ser cuidadosos. Debemos usar el símbolo de integral que incluye el diferencial (la d final, y colocar la variable de integración en ese símbolo de diferencial.

Quizá la mejor manera de entenderlo sea con un ejemplo. En la siguiente figura puedes ver el cálculo de una primitiva de k , hecho de tres maneras con Wiris, que hemos numerado (en rojo) para discutir las.

$$\left[\int k \, dx \rightarrow k \cdot x \quad (1) \right.$$

$$\left[\int k \, dk \rightarrow \frac{1}{2} \cdot k^2 \quad (2) \right.$$

$$\left[\int k \rightarrow \frac{1}{2} \cdot k^2 \quad (3) \right.$$

1. La primera, que es la que necesitamos en este caso, usa el diferencial dx , y en ese caso Wiris interpreta correctamente k como una constante. La primitiva es $F(x) = k \cdot x$.
2. En la segunda, para ilustrar el problema, hemos usado el diferencial dk . Esto hace que Wiris interprete k como una variable, y la primitiva resultante es $\frac{k^2}{2}$. Si usáramos esto cometeríamos un error. Insistimos: para nosotros k es una constante.
3. En el paso anterior, es posible que el lector piense que es un error difícil de cometer, porque tenemos que colocar ex profeso la k en el diferencial. Esta última versión ilustra la forma en que, habitualmente, se comete ese error. Aquí usamos el símbolo de integral sin diferencial, y en ese caso Wiris tiene que decidir, por sí mismo, cual es la variable. Y como sólo aparece la k , Wiris elige esta como variable.

En conclusión, es conveniente que, en Wiris, utilices siempre la integración con diferenciales, y escribas explícitamente la variable con respecto a la que integras.

Wolfram Alpha.

¿Qué sucede en Wolfram Alpha? Algo muy parecido:

1. Cuando escribimos `integrate(k, x)`, indicándole a Wolfram Alpha que la variable de integración es x , obtenemos la respuesta correcta:

$$\int k \, dx = k \cdot x + \text{constant}$$

(Y recuerda que puedes tomar la constante `constant` igual a 0).

2. Si escribimos `integrate(k, k)`, señalando a k como variable de integración es x , obtenemos una respuesta que, en este caos, es un error:

$$\int k \, dk = \frac{k^2}{2} + \text{constant}$$

3. Y por último, y más peligroso, si escribimos `integrate(k)`, y dejamos que sea el programa quien decida, obtenemos de nuevo la respuesta errónea.

GeoGebra.

En GeoGebra hemos empezado por abrir el panel de *Cálculo Simbólico*, y hemos tecleado el comando

$$f(x) := k$$

El símbolo := es la forma de asignar valores a símbolos en este panel de *Cálculo Simbólico*. Una vez hecho esto (y tras pulsar **Entrar**, claro), basta con ejecutar el comando

$$\text{Integral}[f(x), x]$$

en el que, como en los otros casos, hemos tenido que indicar la variable de integración.

Ejercicio 16.

En todos los casos, es conveniente que intentes hacer los ejercicios usando varios de los programas, para cotejar los resultados.

1. En el ejemplo que acabamos de describir con GeoGebra, ejecuta uno tras otro, para comparar los resultados, los comandos

$$\text{Integral}[f(x), k]$$

$$\text{Integral}[f(x)]$$

2. Calcula una primitiva de $f(x) = x \cdot \cos(3x)$
3. Calcula el valor de k para que la función:

$$f(x) = \begin{cases} k \cdot x \cdot \cos(3x) & \text{si } 0 < x < \frac{\pi}{6} \\ 0 & \text{en otro caso.} \end{cases}$$

sea una función de densidad en el intervalo $0 \leq x \leq \frac{\pi}{6}$.

4. Utiliza el ordenador para comprobar las integrales intermedias del Ejemplo 5.5.1 del libro (pág. 165) y también las del Ejemplo 5.5.2 (pág. 168).
5. Usa Wolfram Alpha, GeoGebra o Wiris para resolver la ecuación cuadrática

$$-\frac{2k^2}{3} + 4k - 5 = \frac{1}{2}.$$

que aparece en el Ejemplo 5.5.2.

Soluciones en la página 67. □

3.7. La distribución uniforme con R.

En la Sección 5.4.3 del libro (pág. 163) hemos hablado de las distribuciones uniformes. Una distribución uniforme definida en el intervalo $[a, b]$ tiene soporte en dicho intervalo (es decir, su densidad vale 0 fuera del intervalo) y es constantemente igual a $1/(b - a)$ dentro del intervalo. R nos ofrece varias herramientas para trabajar con densidades uniformes. En concreto, vamos a hablar de las funciones:

punif, dunif, qunif, runif.

Como ves, el sufijo común a todas ellas es **unif**. La primera de ellas, **pbinom**, proporciona la respuesta a la pregunta:

$$P(X \leq k)$$

siendo X una variable uniforme en el intervalo $[a, b]$. Por ejemplo, si X es una variable uniforme definida en el intervalo $[5, 13]$, y queremos saber cual es la probabilidad

$$P(X \leq 11)$$

ejecutaremos en R el siguiente código:

```
punif(11, min = 5, max = 13)
```

```
## [1] 0.75
```

Como ves, las opciones `min` y `max` nos permiten indicarle a R cuál es el intervalo $[a, b]$ en el que se define la distribución uniforme. La función `punif` es, por tanto, la función de distribución de la variable X . Y, por tanto, si queremos calcular la probabilidad de un subintervalo, como en

$$P(8 \leq X \leq 11)$$

basta con tomar la diferencia de los valores de `punif` al final y al principio de ese subintervalo:

```
punif(11, min = 5, max = 13) - punif(8, min = 5, max = 13)

## [1] 0.375

(11 - 8) / (13 - 5)

## [1] 0.375
```

Hemos incluido el cálculo de esa probabilidad mediante la Ecuación 5.18 del libro (pág. 164), para que veas que los resultados coinciden.

Cuando vimos las funciones de R para trabajar con la binomial, hablamos de `dbinom` antes de hablar de `pbinom`. Ahora hemos empezado hablando de `punif`. ¿Qué sucede con `dunif`?

Densidad de variables aleatorias continuas en R. Funciones que empiezan por d.

Los cálculos de probabilidad, en una distribución uniforme, **siempre** se llevan a cabo con `punif`. Al tratarse de una distribución continua, la función `dunif` representa la densidad de probabilidad, y **no es una probabilidad** (hay que integrarla para obtener la probabilidad). En particular, cuando se trata de distribuciones continuas, prácticamente nunca vamos a usar las funciones de R que empiezan por d.

Para insistir en estas ideas vamos a hacer un ejercicio.

Ejercicio 17.

Vamos a definir una variable aleatoria uniforme en el intervalo $\left[\frac{1}{130}, \frac{12}{130}\right]$. Calcula las probabilidades:

1. $P\left(X < \frac{6}{130}\right)$.
2. $P\left(\frac{2}{130} < X < \frac{9}{130}\right)$.
3. Repite los cálculos anteriores cambiando $<$ por \leq . ¿Hay alguna diferencia? ¿Por qué sucede esto?
4. Calcula `dunif(6/130, min=1/130, max=12/130)`. ¿Puede ser ese valor una probabilidad? Calcula también `dunif(20/130, min=1/130, max=12/130)`. Fíjate en que el valor $\frac{20}{130}$ está fuera del intervalo soporte de esta función uniforme.
5. Calcula la probabilidad $P(X > 6/130)$ (y también con \geq en lugar de $>$).

Soluciones en la página 68. □

Como hemos dicho ya antes, R utiliza siempre por defecto probabilidades de la forma $P(X \leq k)$. Si quieres que use desigualdades de la forma $P(X < k)$ puedes usar la opción `lower.tail= FALSE`.

La función `qunif` permite responder a esta pregunta: dada la probabilidad p_0 , ¿cuál es el valor k para el que $P(X \leq k) = p_0$? Es decir, ¿cuál es el cuantil p_0 de X ? Por ejemplo, si X es una distribución uniforme en el intervalo $[-11, 25]$ y queremos calcular su cuantil 0.75 (el tercer cuartil), usaríamos este comando:

```
qunif(0.75, min=-11, max=25)
```

```
## [1] 16
```

Vamos a comprobar el resultado:

Ejercicio 18.

Guarda el resultado del anterior comando en la variable `k` de `R`, y usa `punif` para comprobar que $P(X \leq k) = 0.75$. Solución en la página 69. \square

Jugando con las propiedades de las probabilidades, se puede usar `qunif` para responder a preguntas más generales.

Ejercicio 19.

Sea X una variable aleatoria uniforme en el intervalo $[-2, 8]$. Calcula el valor k para el que se cumple $P(X > k) = \frac{1}{7}$. Comprueba el resultado. Solución en la página 69. \square

Finalmente, la función `runif` sirve para obtener valores aleatorios de una variable aleatoria uniforme. Esta función es extremadamente útil a la hora de hacer simulaciones, porque sirve para poner en práctica esa idea de “lanzar un dardo al azar dentro del intervalo $[a, b]$, de forma que todas las regiones del intervalo sean igual de probables”. Vamos a usar un ejercicio para ver cómo usar esto.

Ejercicio 20.

Sea X una variable aleatoria uniforme en el intervalo $[-5, 5]$.

1. Usa `runif` para generar un vector de $n = 1000$ puntos aleatorios dentro de ese intervalo. Llama `puntos` a ese vector.
2. Calcula, con `punif`, la probabilidad de que uno de esos puntos pertenezca al intervalo $[-1, 3]$.
3. ¿Qué fracción (o porcentaje) de los elementos de `puntos` pertenecen de hecho al intervalo $[-1, 3]$?
4. Repite los apartados anteriores para $n = 10000$ y $n = 1000000$.
5. **Más difícil:** En lugar de GeoGebra, se puede usar `R` para simular el lanzamiento de dardos dentro de un cuadrado de lado 4, y calcular el valor de π , como hemos hecho en el Ejemplo 3.3.3 del libro (pág. 54). Puedes hacer esto así:
 - a) Construyes n valores de x uniformemente distribuidos en el intervalo $[-2, 2]$.
 - b) De la misma forma, construyes n valores de y uniformemente distribuidos en el intervalo $[-2, 2]$. Estos valores de y son, desde luego, independientes de los de x .
 - c) Al emparejar cada valor de x con el correspondiente valor de y el punto (x, y) resultante es un punto al azar en el cuadrado de lado 4 centrado en el origen.
 - d) Para saber si un punto de coordenadas (x, y) ha caído en el interior del círculo de radio 1, basta con ver si se cumple la condición

$$x^2 + y^2 < 1.$$

- e) Una vez que sepas la fracción de puntos que han caído en el círculo, multiplica esa fracción por 16 (el área del cuadrado), para estimar qué fracción del área del cuadrado corresponde al área del círculo de radio 1. Eso te permitirá obtener un valor aproximado de π .

La ventaja de `R`, frente a GeoGebra, es que `R` puede usar un número enorme de puntos sin problemas. Usa este esquema, y un número muy grande de puntos (del orden de millones) para estimar el valor de π . En el código de la solución veremos, además, las instrucciones necesarias para dibujar cada uno de los pasos anteriores.

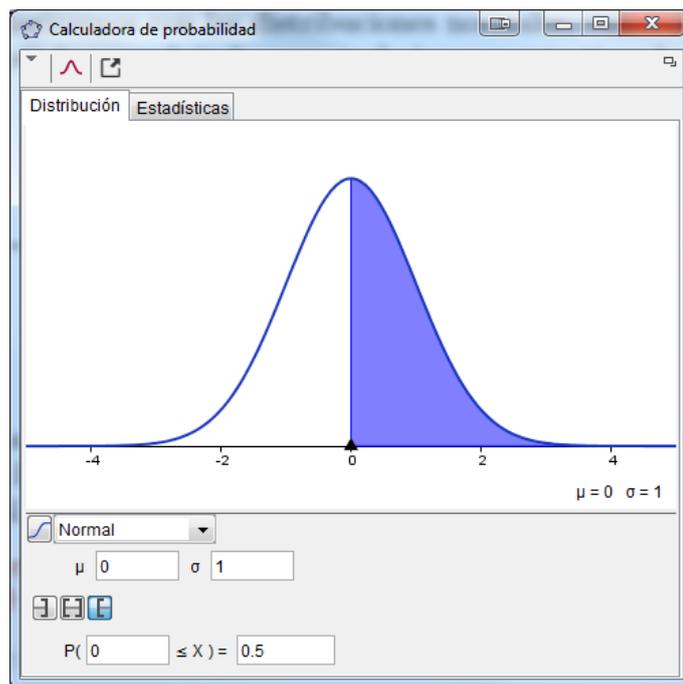
Soluciones en la página 70. \square

4. La distribución normal.

En esta sección nos vamos a ocupar de la más importante de todas las distribuciones continuas, la normal. Veremos cómo usar los programas de ordenador que conocemos (GeoGebra, Wolfram Alpha y, desde luego, R) para trabajar con variables aleatorias de tipo normal.

4.1. La familia de las curvas normales con GeoGebra.

Para empezar a familiarizarnos con las distribuciones normales, vamos a comenzar conociendo las herramientas que GeoGebra nos brinda para trabajar con este tipo de distribuciones. En primer lugar, vamos a volver a abrir la *Calculadora de Probabilidades* que ya conocimos al tratar con la distribución binomial en la Sección 1 de este Tutorial. Al abrirla, GeoGebra nos muestra precisamente la distribución normal.



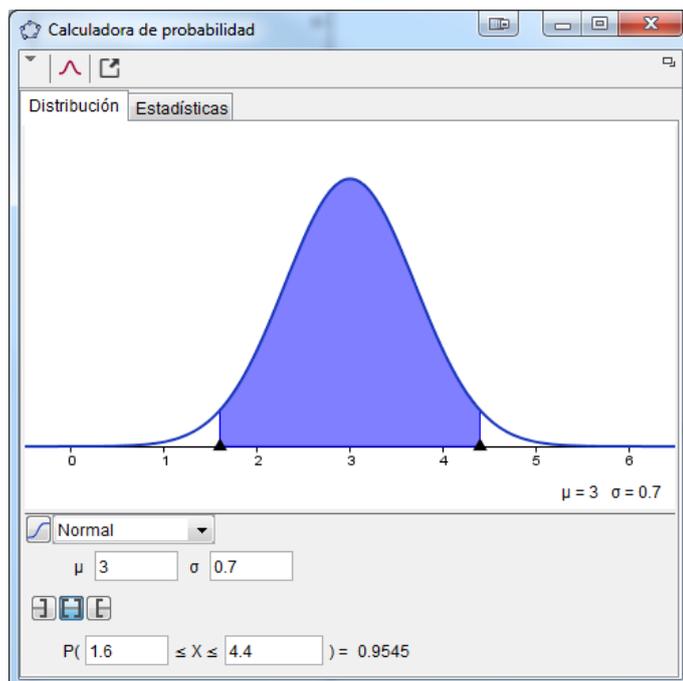
Para ver cómo usar la *Calculadora de Probabilidades*, vamos a considerar una variable aleatoria X de tipo $N(3, 0.7)$ (esto es, con media $\mu = 3$ y desviación típica $\sigma = 0.7$) y vamos a calcular la probabilidad:

$$P(1.6 < X < 4.4)$$

Para hacerlo, introduce los valores de μ y σ en los campos adecuados de la *Calculadora de Probabilidades*. Después, y puesto que se trata de un intervalo, haz clic con el ratón en el icono central del grupo de iconos



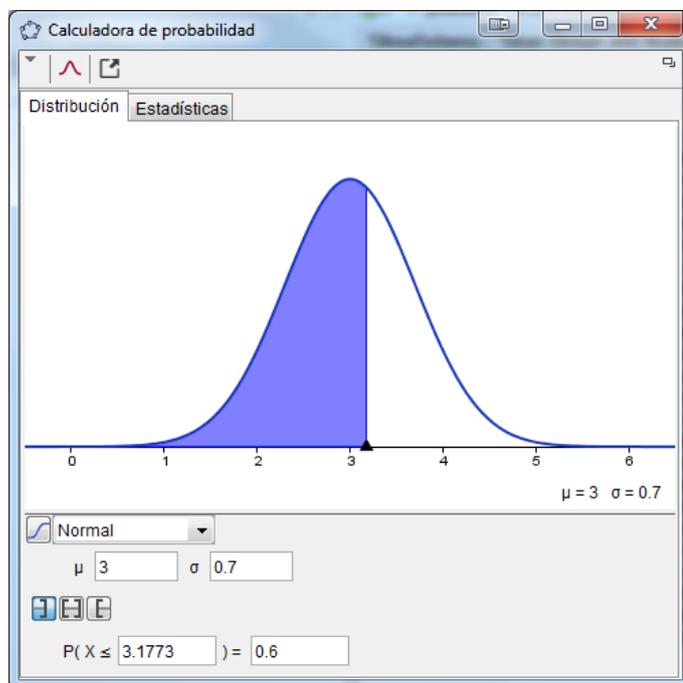
que ya discutimos al hablar de la binomial. Al seleccionar ese icono la parte inferior de la *Calculadora de Probabilidades* cambia para adaptarse al tipo de pregunta que queremos formular. Asegúrate de introducir los extremos del intervalo 1.6 y 4.4 en los campos adecuados, para que la ventana de la *Calculadora de Probabilidades* tenga este aspecto:



Como ves, la probabilidad que buscábamos es, aproximadamente, 0.9545. La solución, además, se ilustra gráficamente. La *Calculadora de Probabilidades* también se puede utilizar para calcular cuantiles. Por ejemplo, en esa misma variable de tipo $N(3, 0.7)$ vamos a buscar el valor k tal que:

$$P(X \leq k) = 0.60$$

Es decir, el percentil 60 de esta variable. Para ello, selecciona el icono izquierdo del grupo , y teclea 0.60 en el campo situado a la derecha del =, como se ve en la Figura.



Tras pulsar **Entrar**, GeoGebra te mostrará (dentro del símbolo de probabilidad) que el valor de k es, aproximadamente, 3.1773 (y, de nuevo, la solución se ilustra gráficamente). Fíjate en que, al ser 0.6 ligeramente mayor que $1/2$, la probabilidad incluye toda la mitad izquierda del área bajo la curva, y una pequeña parte de la mitad derecha del área. Es una buena idea acostumbrarse a pensar geoméricamente en las probabilidades que estamos calculando, tratando siempre de traducirlas en términos de áreas. Vamos a proponerte una serie de ejercicios relacionados con el cálculo de probabilidades en las distribuciones normales. Es un tipo de ejercicios que nos resultará muy útil más adelante. Insistimos en que trates de buscar siempre el sentido geométrico de estos cálculos.

Ejercicio 21.

En los siguientes ejercicios sea X una variable aleatoria de tipo $N(5, 3)$. Calcula las probabilidades y valores que se indican.

1. $P(X > 6)$, $P(X > 7)$ y, finalmente, $P(X > 10)$. ¿Qué observas?
2. $P(X < 4)$, $P(X < 3)$ y, finalmente, $P(X < 0)$. ¿Ves alguna relación con los valores del apartado anterior?
3. $P(4.5 < x < 5.5)$, $P(2 < X < 8)$ y, finalmente, $P(0 < X < 10)$.
4. Los valores k_1 y k_2 tales que $P(X < k_1) = 0.90$ y $P(X < k_2) = 0.95$.
5. Los valores k_1 y k_2 tales que $P(X > k_1) = 0.1$ y $P(X > k_2) = 0.05$. ¿Ves alguna relación con los valores del apartado anterior?

Ahora, con la misma variable X , responde a estas preguntas, pero sin usar el ordenador:

6. El valor $P(X < 7)$ ¿es mayor o menor que $\frac{1}{2}$?
7. El valor $P(X > 8)$ ¿es mayor o menor que $\frac{1}{2}$?
8. ¿Cuál de estos dos valores es más grande, $P(X > 4)$ o $P(X < 5.5)$?
9. El valor k tal que $P(X > k) = 0.6$ ¿es mayor o menor que 5?
10. El valor k tal que $P(X < k) = 0.1$ ¿es mayor o menor que 5?

Para la última parte, vuelve a la ventana principal de GeoGebra, ve al menú Opciones, y en Redondeo asegúrate de seleccionar 5 cifras decimales.

11. Para comprobar los resultados del Ejemplo 5.2.1 del libro (pág. 143), ejecuta en la Línea de Entrada de GeoGebra, uno tras otro, los tres comandos siguientes:

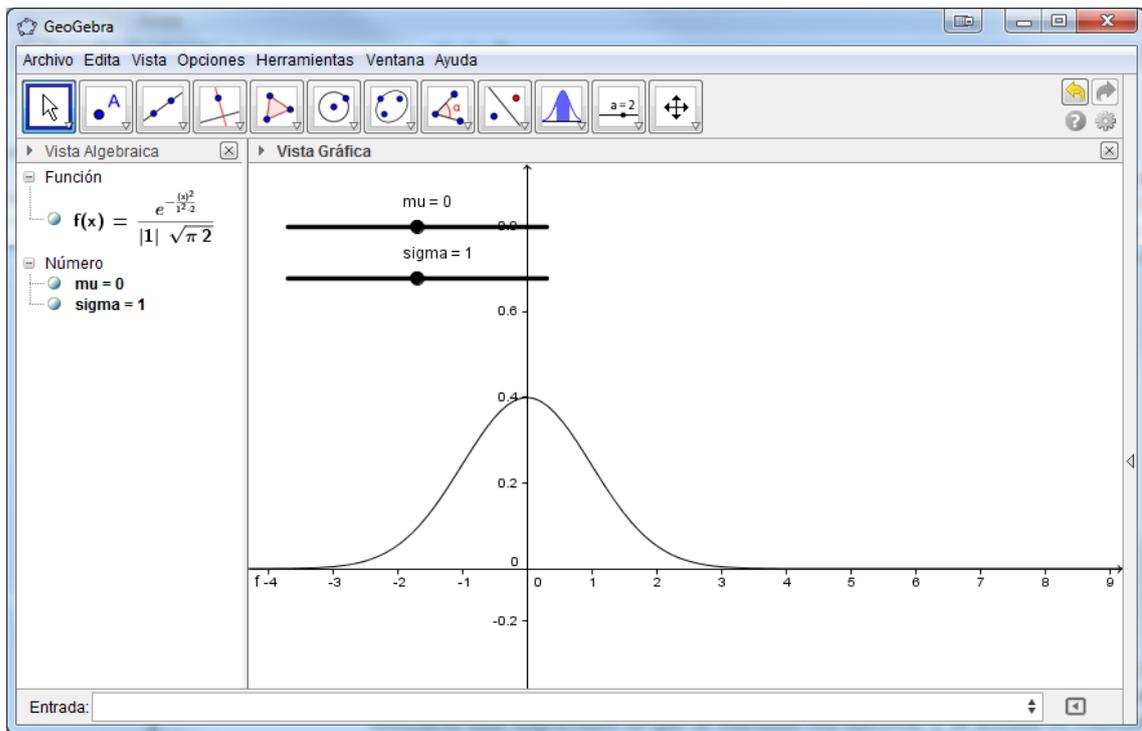
```
DistribuciónBinomial[100,1/3,30,false]
f(x) = Normal[100 * (1/3), sqrt(100 * (1/3) * (2/3)), x]
f(30)
```

Soluciones en la página 74. Para el último apartado las soluciones aparecen en el Ejemplo 5.2.1 del libro. □

Como hicimos con la binomial, una forma de familiarizarse con las curvas normales es utilizar las capacidades dinámicas de GeoGebra para variar (mediante deslizadores) los valores de μ y σ , y ver el efecto que tienen sobre la forma y posición de la curva. Para hacer esto, abre una nueva ventana de GeoGebra y, desde la Línea de entrada, ejecuta consecutivamente estos comandos:

```
mu = Deslizador[-5, 5, 0.05, 0, 200]
sigma = Deslizador[0, 2, 0.05, 0, 200]
sigma = 1
Normal[mu, sigma, x] RazónEjes [20,3]
```

Puedes usar los nombres `mu` y `sigma`, como hemos hecho nosotros, o puedes usar los símbolos μ y σ , con el panel de caracteres de GeoGebra, que aparece a la derecha de la Línea de entrada. El comando `sigma = 1` nos sirve para fijar el valor inicial de `sigma`, y comenzar con la normal $N(0, 1)$. El comando `RazónEjes [20,3]` ajusta la relación de escalas de los dos ejes para que la visualización resulte más cómoda. Tras ejecutar todos estos comandos, tu ventana de GeoGebra tendrá un aspecto similar a este:



Fíjate en la ecuación de la curva normal, que aparece en el panel de *Vista Algebraica*. Ahora puedes empezar a experimentar con los deslizadores de μ y σ . ¿Qué sucede al cambiar μ o σ ? ¿Cambia la forma o posición de la curva?

En cualquier caso, es importante que recuerdes que, por mucho que cambien la forma o la posición de la curva, el área total bajo esa curva siempre es igual a 1. Para insistir en esta idea, y en cómo se distribuye la probabilidad con las curvas normales, te hemos preparado otro fichero GeoGebra, muy parecido a lo que acabamos de ver:

[Tut05-CurvasNormalesAreas.ggb](#)

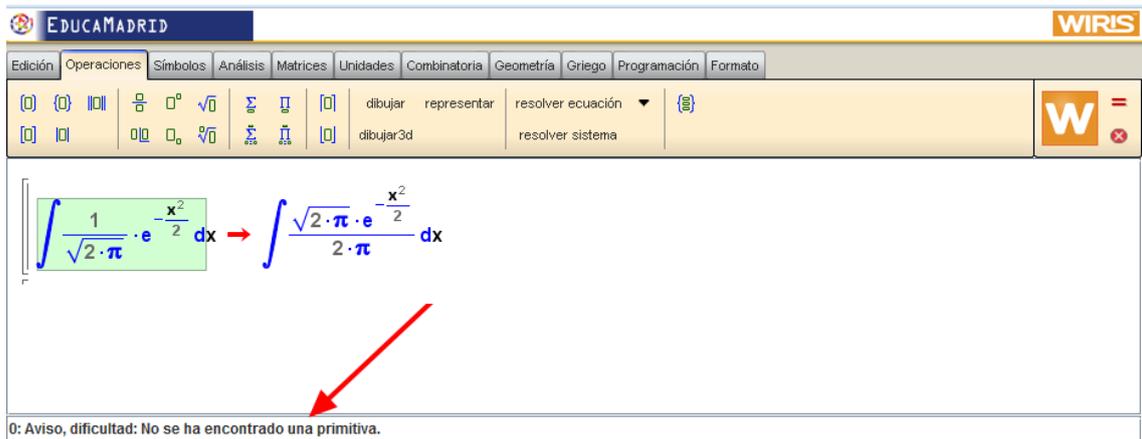
pero en el que puedes hacer clic en la casilla rotulada “*Muestra áreas*”, para ver en acción lo que se conoce como *Regla 68-95-99 para las distribuciones normales* (ver Ecuación 5.22 en el libro, pág. 175)

4.2. Tratando de calcular una primitiva de la curva normal.

En la discusión de la página 176 del libro hemos dicho que no es posible encontrar una primitiva de la curva normal. En un intento de hacer esta idea más cercana (ya que es bastante abstracta, a nuestro juicio), vamos a ver lo que sucede cuando le pedimos a un programa de ordenador que trate de calcular esa primitiva. En concreto, para simplificar pero sin merma de generalidad, vamos a quedarnos en el caso de la normal $N(0, 1)$, que es la que tiene la ecuación más sencilla de todas (Ecuación 5.24 del libro, pág. 177):

$$f_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

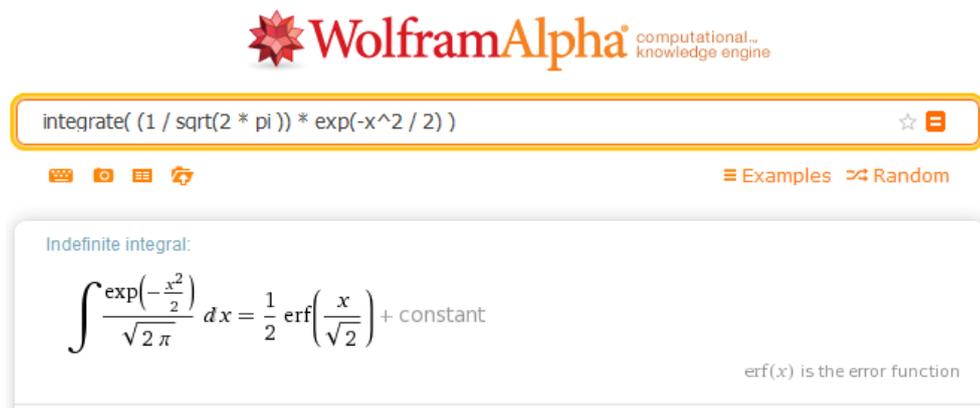
Tratemos de calcular una primitiva usando Wiris. La siguiente figura muestra el resultado, y la flecha roja señala el mensaje con el que Wiris nos avisa de la dificultad con la que hemos tropezado.



En Wolfram Alpha sucede algo ligeramente distinto. Escribe este comando en el programa:

```
integrate( ( 1 / sqrt(2 * pi )) * exp(-x^2 / 2) )
```

y verás que, en su respuesta, Wolfram Alpha usa una función llamada *erf*, del inglés *error function*.



¿Qué es esta *función error erf*? Pues en el fondo, es simplemente un nombre o símbolo, que el programa utiliza para referirse a la primitiva

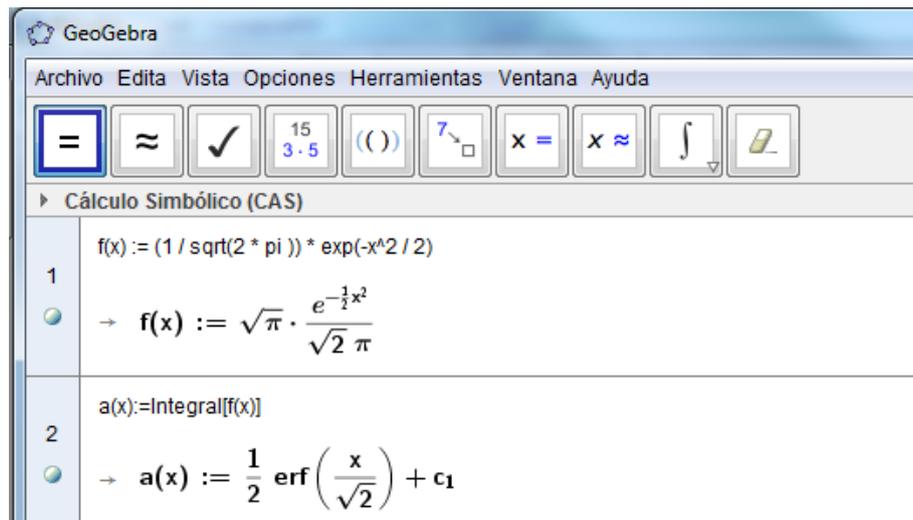
$$\int e^{-x^2} dx$$

porque Wolfram Alpha sabe que esta función no tiene una primitiva elemental. En el fondo, lo que Wolfram Alpha ha hecho ha sido avisarnos de que se ha dado cuenta de lo que estamos tratando de hacer, y usar el nombre *erf* como abreviatura de “una primitiva de $e^{-x^2/2}$ ”. Es un mensaje tranquilizador, porque si necesitamos calcular valores de esa función *erf*, Wolfram Alpha nos los proporcionará sin dificultad. Por ejemplo, prueba a ejecutar

```
erf(2)
```

y el programa te responderá que ese valor es, aproximadamente, 0.9953. No queremos ponernos demasiado profundos con este tema, pero en realidad, la función **erf** es una función tan “extraña” como puedan serlo la función seno, o el logaritmo. El hecho es que en la formación matemática elemental nos han hablado de esas funciones, como las trigonométricas y el logaritmo, y nos hemos acostumbrado a verlas como teclas de la calculadora, etc. Además, las funciones trigonométricas tienen interpretaciones geométricas sencillas. Y por eso las llamamos elementales. Pero el coseno tiene tanto de elemental como pueda tenerlo *erf*. Para que el lector nos entienda, ¿cómo se calcula $\cos(0.32)$ “a mano”? Al final, todos recurrimos a la calculadora o el ordenador para responder (salvo los ingenieros encargados de diseñar la calculadora, y los matemáticos que tienen que diseñar el método que usará la calculadora para hacer ese cálculo). Es en ese sentido en el que decimos que puedes pensar en *erf* como una “nueva tecla de la calculadora”, y dejar los detalles para los matemáticos e ingenieros encargados del asunto.

Por cierto, GeoGebra, en su panel de *Cálculo Simbólico*, también usa la función *erf* para contestar, como ilustra esta figura:



5. La distribución normal en R.

Al tratar con la binomial en R conocimos las funciones `dbinom`, `pbinom`, `qbinom` y `rbinom`. En la distribución uniforme, las correspondientes funciones fueron `dunif`, `punif`, `qunif` y `runif`. No es ninguna sorpresa, por tanto, que ahora, para trabajar con distribuciones normales, sea el turno de las funciones

`dnorm`, `pnorm`, `qnorm`, y `rnorm`.

Veamos por turno el significado, y la forma de usar de cada una de estas funciones. Empezando por la primera, y de forma similar a la discusión que tuvimos en el caso de `dunif`, hay que recordar que la normal es una distribución continua. Por esa razón, **en este curso apenas usaremos la función `dnorm`**.

5.1. La función `pnorm`

Como ya hemos dicho, le vamos a dar poco o ningún uso a la función `dnorm`. En cambio, la función `pnorm` nos va a resultar extremadamente útil. Si tenemos una variable X de tipo $N(\mu, \sigma)$ y queremos calcular el valor

$$P(X \leq k)$$

(es decir, lo que llamamos la cola izquierda de la distribución normal en el punto k), lo podemos hacer en R con estas instrucciones:

```
mu = 3
sigma = 2
k = 4
pnorm(k, mean=mu, sd=sigma)

## [1] 0.6915
```

Como ves, las opciones `mean` y `sd` le indican a R cuáles son los parámetros μ y σ , respectivamente, de la normal (el nombre `sd` es desafortunado, porque nos hace pensar en la cuasidesviación típica).

Antes de seguir adelante, un par de observaciones.

- Queremos insistir en algo que ya vimos en el caso de la distribución uniforme y que es esencial entender. Como la distribución normal es continua, siempre se cumple esta igualdad:

$$P(X < k) = P(X \leq k).$$

En las distribuciones continuas no hay diferencia entre desigualdades estrictas o no estrictas (recuerda que la probabilidad de un punto es 0 en estos casos). Esto contrasta claramente con lo que vimos para la binomial, donde la diferencia entre ambos tipos de desigualdades es muy importante, e ignorarla produce siempre errores.

- Aparte de esto, también es muy importante recordar que, como ya anunciamos con la binomial, R siempre trabaja por defecto con la cola izquierda (lo mejor es pensar en \leq). Es decir, que por defecto R usa desigualdades de la forma $P(X \leq k)$, sea cual sea el tipo de la variable X (binomial, normal, uniforme, etc.)

Si tenemos en cuenta estas dos ideas, podemos calcular cualquier valor de probabilidad de la distribución normal combinando la función `pnorm` con las propiedades básicas de la probabilidad. Por ejemplo, si X es de tipo $N(10, 2)$ y queremos calcular la probabilidad (de una cola izquierda)

$$P(X < 10.5)$$

usaríamos

```
pnorm(10.5, mean=10, sd=2)
## [1] 0.5987
```

Si lo que queremos calcular es (una cola derecha)

$$P(X > 11)$$

usaríamos

```
1-pnorm(11, mean=10, sd=2)
## [1] 0.3085
```

Y si queremos calcular la probabilidad de un intervalo, como

$$P(7 < X < 12)$$

usaríamos

```
pnorm(12, mean=10, sd=2) - pnorm(7, mean=10, sd=2)
## [1] 0.7745
```

Estos tres ejemplos cubren, en realidad, todas las situaciones que se presentan en la práctica, cuando de trata de calcular probabilidades para la distribución normal. Vamos a practicar el uso de `pnorm` en un ejercicio.

Ejercicio 22.

1. Repite, usando `pnorm`, los apartados 1 a 3 del Ejercicio 21 (pág. 35)
2. Comprueba, usando `pnorm`, los valores de probabilidad que aparecen en el Ejemplo 5.6.2 del libro (pág. 181), tanto usando la corrección de continuidad como sin usarla. Es decir, dada una variable aleatoria normal $Y \sim N\left(7, \sqrt{\frac{14}{3}}\right)$, calcula las probabilidades:

$$P(5 \leq Y \leq 9)$$

y

$$P(4.5 \leq Y \leq 9.5).$$

3. Elige (por ejemplo, usando `sample` o `runif`) dos valores cualesquiera de μ y σ , y usa `pnorm` para comprobar la Regla 68-95-99 (Ecuación 5.22 del libro, pág. 175).

Soluciones en la página 75. □

5.2. La función qnorm

La función `qnorm`, como `qbinom` y `qunif`, sirve para resolver problemas inversos de probabilidad, o problemas de cuantiles. Por ejemplo, en una distribución normal de tipo $N(5, 2)$, ¿cuál es el valor k para el que se cumple esta ecuación?

$$P(X \leq k) = \frac{1}{3}.$$

Para este ejemplo anterior la respuesta se obtiene ejecutando:

```
(k = qnorm(1/3, mean=5, sd=2))  
## [1] 4.139
```

Como es de esperar, podemos verificarlo con `pnorm`

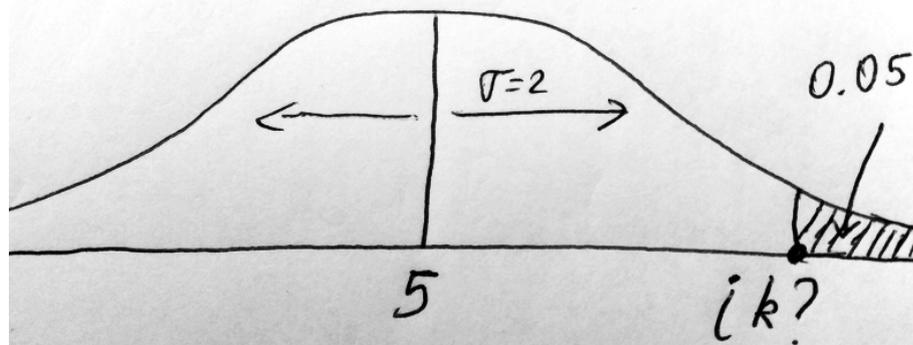
```
pnorm(k, mean=5, sd=2)  
## [1] 0.3333
```

que es aproximadamente igual a $1/3$. Cuando, en el Capítulo 6 del libro empecemos a hacer Inferencia, veremos que el tipo de preguntas que `qnorm` son extremadamente importantes para calcular, por ejemplo, intervalos de confianza basados en la distribución normal (los primeros que encontraremos).

Para otros problemas inversos de probabilidad tendremos que usar los trucos que hemos visto en casos anteriores: modificar un poco la llamada a la función `qnorm`, o usar la opción `lower.tail`. Por ejemplo, con la misma variable $X \sim N(5, 2)$, vamos a tratar de averiguar cuál es el valor k para el que se cumple (es una cola derecha)

$$P(X > k) = 0.05$$

Cuando te enfrentes con un problema como este (y, créeme, a lo largo del curso eso va a suceder muchas, muchas veces) te recomiendo encarecidamente que busques un papel y trates de hacer un esquema, por rudimentario que sea, de la situación y de lo que estás tratando de calcular. Para que veas que el dibujo puede ser realmente básico, ahí tienes el que yo me he hecho para esta situación:



Como ves, no se trata de ser preciso, sino de capturar los ingredientes básicos de la situación. En este ejemplo, indicamos que la media μ se sitúa en 5 (siendo $\sigma = 2$), y que lo que estamos buscando es el valor de k que deja a su derecha una probabilidad (el área sombreada) igual a 0.05. Así pues, el resto del área, lo que queda a la izquierda de k , vale 0.95. Esta figura nos permite ver con claridad que lo que tenemos que pedirle a R es el valor:

```
(k = qnorm(0.95, mean=5, sd=2))  
## [1] 8.29
```

Si deseas figuras más precisas, recuerda que puedes utilizar la *Calculadora de Probabilidades* de GeoGebra. Pero en la mayoría de los casos, un papel, un bolígrafo y un poco de concentración serán tus mejores aliados. Para que vayas afinando la puntería, aquí tienes algunos ejercicios.

Ejercicio 23.

Intenta, en todos los casos, hacer previamente un dibujo básico de la situación.

1. Repite, usando `qnorm`, los apartados 4 y 5 del Ejercicio 21 (pág. 35).
2. Sea X una variable normal de tipo $N(-2, 1/4)$. Calcula el valor de k tal que $P(X < k) = 0.85$.
3. Para la misma variable del apartado anterior, calcula el valor de k para el que $P(X > k) = 0.99$.
4. Más divertido y, a la vez, muy importante. Busca, para la misma variable X , el valor de k tal que $P(-2 - k < X < -2 + k) = 0.95$. Es realmente útil que trates de visualizar la situación. Insistimos, este apartado es **muy importante**. Esfuérate en entender el resultado, será una inversión rentable en próximos capítulos.

Soluciones en la página 76. □

La normal estándar Z

De entre todas las distribuciones normales, la más simple, y a la vez más destacada, es la que tiene media 0 y desviación típica 1, a la que llamamos normal estándar y que representamos con la letra Z . La normal estándar ocupa un lugar destacado en la Estadística, y en R no podía suceder otra cosa. De hecho, si no le damos ninguna información sobre la media y la desviación típica, R siempre asume que la normal de la que hablamos es la estándar. Así pues, si escribimos

```
pnorm(1.5)
## [1] 0.9332
```

entonces R nos responde asumiendo que estamos interesados en conocer el valor de

$P(Z < 1.5)$

donde, insistimos, Z representa a una normal estándar de tipo $N(0, 1)$. Las funciones `dnorm`, `pnorm`, `qnorm` y `rnorm` aplican todas el mismo principio: una normal, por defecto, es la normal estándar Z , salvo que indiquemos su media y desviación típica explícitamente.

5.3. La función `rnorm`

Ya vimos como usar la función `rbinom` para generar valores aleatorios de una distribución binomial. Así que, siguiendo el convenio de notación de R, no debería sorprendernos que la función `rnorm` nos proporcione valores aleatorios de una distribución normal. Por ejemplo

```
rnorm(50, mean=100, sd=5)
## [1] 97.66 102.85 102.51 106.13 105.23 105.31 102.47 101.94 96.64 100.54
## [11] 99.77 113.09 94.33 98.48 101.06 99.77 98.51 98.74 105.98 96.47
## [21] 99.47 102.12 88.27 102.66 99.29 98.64 95.60 98.76 103.79 103.46
## [31] 95.76 93.26 91.28 95.87 105.23 91.66 99.51 99.50 100.66 99.94
## [41] 92.88 100.45 107.12 102.86 106.95 105.74 107.78 98.31 101.58 110.86
```

nos proporciona 50 valores aleatorios de una distribución de tipo $N(100, 5)$. Las variables normales son esenciales en Estadística, así que la posibilidad de generar valores (pseudo)aleatorios que siguen una distribución normal dada nos abre la puerta a muchas simulaciones y experimentos interesantes.

Ejercicio 24.

1. En el Ejemplo 5.6.1 del libro (pág. 178) se afirma que, si $X \sim N(400, 15)$, entonces

$$P(380 \leq X \leq 420) \approx 0.82.$$

El objetivo de este ejercicio es comprobar ese resultado mediante una simulación. Para ello:

- Vamos a generar, usando `rnorm`, un vector \mathbf{X} con muchos valores de esa distribución normal (muchos quiere decir miles o decenas de miles, ¡no seamos tímidos!).
 - Identifica los elementos del vector \mathbf{X} que pertenecen al intervalo $(380, 420)$.
 - Calcula qué fracción del total representan esos elementos del vector \mathbf{X} . Este valor es una estimación de la probabilidad $P(380 \leq X \leq 420)$.
 - Finalmente, usa `pnorm` para calcular un valor aproximado (no simulado, pero sin duda más exacto) de esa probabilidad.
2. Otra simulación interesante tiene que ver con las propiedades de la suma de distribuciones normales que se discuten en la página 178 del libro. Allí hemos dicho que si

$$X_1 \sim N(\mu_1, \sigma_1) \quad \text{y} \quad X_2 \sim N(\mu_2, \sigma_2),$$

son variables normales independientes, su suma es de nuevo una variable normal de tipo

$$N\left(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right).$$

Para comprobar esto “experimentalmente”, haremos lo siguiente:

- Vamos a tomar como ejemplo $X_1 \sim N(15, 3)$ y $X_2 \sim N(30, 4)$. Usando `rnorm`, genera dos vectores en R, llamados $\mathbf{X1}$ y $\mathbf{X2}$, con $n = 100000$ (o más) elementos cada uno, correspondientes a esas dos distribuciones normales.
- Calcula el vector suma \mathbf{X} .
- Calcula su media y cuasidesviación típica muestral. ¿Son los valores que esperabas? Dibuja el histograma de X para comprobar que tiene el aspecto de una distribución normal.
- Para la última simulación de este ejercicio vamos a fijarnos en la regla 68-95-99 que aparece en la página 175 del libro. Tu objetivo es diseñar una simulación para comprobar esa regla en una variable aleatoria normal cualquiera. Puedes, de hecho, elegir al azar la media y desviación típica de la normal como parte de la simulación.

Soluciones en la página 77. □

5.3.1. Tipificación en R.

Para tipificar valores, R pone a nuestra disposición la función `scale`. Esta función usa como argumentos un vector de datos X , una media μ y una desviación típica σ , y nos devuelve como resultado el vector que se obtiene al aplicar la transformación:

$$\frac{X - \mu}{\sigma}.$$

Fíjate en que no estamos diciendo que X tenga que ser un vector de datos de una distribución normal, ni nada parecido. La operación es puramente aritmética (restar y dividir). Te recomendamos que leas el fichero de ayuda de `scale` para entender bien cómo se usa (recuerda el tabulador de RStudio para llegar hasta esa ayuda).

6. Definición de funciones e integración en R

Opcional: esta sección puede omitirse en una primera lectura.

En este Tutorial hemos visto (algunas de) las herramientas que R nos ofrece para trabajar con la distribución binomial (que es discreta), y con la distribución normal (que es continua). Se trata en ambos casos, de distribuciones muy importantes, con nombre propio, por así decirlo, y a las que

R conoce bien. Por eso existen funciones específicas como `dbinom`, `pnorm`, etc., para trabajar con estas distribuciones en concreto.

Por otra parte, en Tutorial04 hemos visto como utilizar R para tratar con variables aleatorias discretas genéricas (que no tienen nombre propio, al contrario que la binomial), definidas mediante una tabla de densidad de probabilidad. Y la pregunta es evidente: ¿qué ocurre con las variables aleatorias genéricas de tipo continuo? ¿Cómo las podemos manejar desde R? Para responder a esas preguntas, vamos a aprender dos cosas:

- En primer lugar, aprenderemos a definir funciones en R.
- A continuación, veremos como calcular sus integrales en un intervalo de valores. El cálculo, tratándose de R, siempre será *numérico*, es decir, aproximado.

6.1. Funciones en R

Empecemos con la definición de una función en R. Queremos subrayar desde el principio que, aunque aquí las vamos a usar como funciones de densidad, las funciones se pueden usar para muchas otras cosas en R (veremos ejemplos de esto en el futuro). La definición de funciones (de una o más variables) es una de las características que hacen de R un programa tan flexible y potente.

Vamos a empezar con un ejemplo realmente elemental. Definiremos una función, llamada **cuadrado**, que simplemente calculará el cuadrado de un número dado. Para hacerlo, usamos este código:

```
cuadrado = function(x){ x^2 }
```

Como ves, se usa `function` para definir la función, y el resultado se asigna al nombre **cuadrado** que hemos elegido para la función. Entre paréntesis aparece el nombre de la variable (o variables) de la que depende la función, que en este ejemplo es `x`. Y las llaves `{ }` delimitan lo que llamaremos el *cuerpo de la función*, que contiene las operaciones que hay que hacer con `x` para obtener el valor de la función. En el ejemplo, el cuerpo de la función contiene simplemente la operación `x^2`. Una vez definida, vamos a usar la función para calcular el cuadrado de un número, por ejemplo de 5.

```
cuadrado(5)
```

```
## [1] 25
```

Es más que probable que estés pensando que, si lo que queríamos era el cuadrado de 5, resultaba mucho más fácil escribir `5^2` y zanjar el problema. Desde luego, es cierto. Pero esto es sólo el primer ejemplo. Las funciones de R pueden ser mucho más complicadas. En el Ejemplo 5.4.5 del libro (pág. 157), que ya hemos usado antes en este tutorial, teníamos la función de densidad de una variable aleatoria continua con soporte en el intervalo $[0, 1]$, que era:

$$f(x) = \begin{cases} 6 \cdot x \cdot (1 - x) & \text{si } 0 \leq x \leq 1, \\ 0 & \text{en otro caso.} \end{cases} \quad (1)$$

Para definir esta función en R usamos este comando (hemos elegido `f` como nombre de la función, pero podríamos usar cualquier otro):

```
f = function(x){
  if((x>0) & (x<1)){
    6 * x * (1 - x)
  } else {
    0
  }
}
```

En este caso, el cuerpo de la función contiene una estructura `if/else` completa, como las que vimos en la Sección 3 del Tutorial04. Por lo demás, el uso de la función es igual de sencillo que antes. Para calcular su valor cuando $x = \frac{1}{2}$ (dentro del soporte), hacemos:

```
f(1/2)
```

```
## [1] 1.5
```

Y si quieres calcular un valor fuera del soporte, por ejemplo $f(5)$, entonces entra en acción la otra rama del `if/else` y se obtiene:

```
f(5)
```

```
## [1] 0
```

como deseábamos.

Ejercicio 25.

Usa R para definir la función de densidad

$$f(x) = \frac{1}{\pi(1+x^2)}$$

del Ejemplo 5.4.1 del libro (pág. 150). Llámala f_2 , porque vamos a seguir usando la función $f(x)$ que hemos definido antes. Solución en la página 80. \square

Para trabajar con las funciones de R es esencial entender estas dos ideas:

- Una vez definida una función, podemos usarla como las otras funciones de R, que hemos ido aprendiendo. Podemos, por ejemplo, aplicar la función a un vector. Para ver esto, vamos a usar la función cuadrado que hemos definido antes para calcular el cuadrado de los números de 1 a 100:

```
cuadrado(1:100)
```

```
## [1] 1 4 9 16 25 36 49 64 81 100 121
## [12] 144 169 196 225 256 289 324 361 400 441 484
## [23] 529 576 625 676 729 784 841 900 961 1024 1089
## [34] 1156 1225 1296 1369 1444 1521 1600 1681 1764 1849 1936
## [45] 2025 2116 2209 2304 2401 2500 2601 2704 2809 2916 3025
## [56] 3136 3249 3364 3481 3600 3721 3844 3969 4096 4225 4356
## [67] 4489 4624 4761 4900 5041 5184 5329 5476 5625 5776 5929
## [78] 6084 6241 6400 6561 6724 6889 7056 7225 7396 7569 7744
## [89] 7921 8100 8281 8464 8649 8836 9025 9216 9409 9604 9801
## [100] 10000
```

Aunque para que esto funcione, la operación que define la función tiene que ser una operación “vectorializable”. ¿Qué queremos decir con esa palabreja? En la Sección 6.2 daremos más detalles.

- Las variables que usamos al definir las funciones son, en algún sentido, mudas, como las variables de las integrales. Es decir, que aunque hayamos definido `cuadrado` usando el nombre `x` para la variable, el siguiente fragmento de código funciona sin problemas.

```
y = 3
cuadrado(y)
```

```
## [1] 9
```

Ejercicio 26.

En este ejercicio queremos que descubras lo que sucede si tratamos de calcular `f(1:100)` para la función `f` que hemos definido antes, la que corresponde a la definición:

$$f(x) = \begin{cases} 6 \cdot x \cdot (1-x) & \text{si } 0 \leq x \leq 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Verás que el resultado no es lo que se esperaba. La manera de salir de este aparente atolladero es usando la función `ifelse`, de la que hablaremos a continuación. Solución en la página 80. \square

6.2. La función `ifelse` para condicionales vectoriales.

Como hemos dicho, si se trata de aplicar una estructura condicional `if/else` a un vector, R sólo usa el primer elemento del vector (y nos avisa con un mensaje de tipo `warning`, que es una *advertencia*, no un error). Veámoslo, mostrando primero el caso en el que `x` es un número, y después cuando es un vector de 10 números:

```
x = 5
if(x > 4){
  x^2
} else {
  0
}

## [1] 25

x = 1:10
if(x > 4){
  x^2
} else {
  0
}

## Warning in if (x > 4) {: la condición tiene longitud > 1 y sólo el primer elemento
será usado

## [1] 0
```

En la segunda parte, cuando `x` es el vector `1:50`, R usa sólo el primer elemento (igual a 1) para las operaciones de la estructura `if/else` (y lanza la mencionada advertencia). ¿Cómo podemos conseguir que esa condición se aplique a todos los elementos del vector? Mediante la función `ifelse`. Por ejemplo, para la segunda parte del código anterior usaríamos:

```
x = 1:10
ifelse(x > 4, x^2, 0)

## [1] 0 0 0 0 25 36 49 64 81 100
```

Como ves, la función tiene tres argumentos. El primero es la condición booleana (con resultado `TRUE/FALSE`), que ahora se evalúa para todos y cada uno de los elementos del vector `x`. El segundo argumento es el resultado en los casos en los que la condición resulta `TRUE`, y el tercer argumento es el resultado cuando la condición resulta ser `FALSE`.

Ejercicio 27.

Reescribe la función `f` usando `ifelse`, para que sea posible aplicarla a vectores. Solución en la página 80. □

6.3. Integración numérica con R. Media y varianza de variables aleatorias continuas genéricas.

Ahora que ya sabemos definir una función de densidad en R, vamos a pasar a la segunda parte del plan. Aprenderemos a calcular (siempre aproximadamente) la integral de la función en un intervalo. Y como aplicación de esto, vamos a calcular la media y desviación típica de la variable aleatoria definida por esa función de densidad.

La integración se lleva a cabo mediante el comando `integrate` de R. Debemos indicar el nombre de la función, y los extremos del intervalo de integración, mediante las opciones `lower` y `upper`. Por ejemplo, usando la función `f` que hemos definido en el apartado anterior, podemos repetir el

segundo apartado del Ejercicio 13 (pág. 25), en el que calculábamos la integral

$$\int_{\frac{1}{2}}^{\frac{3}{4}} 6 \cdot (x - x^2) dx.$$

Este cálculo se puede realizar en R mediante:

```
integrate( f, lower=1/2, upper=3/4)
## 0.3438 with absolute error < 3.8e-15
```

y el resultado es el mismo que obtuvimos en el Ejercicio 13, aunque allí la respuesta era simbólica ($\frac{11}{32} \approx 0.3438$). Como ves, R además nos da información sobre la calidad de la aproximación numérica a la integral. Eso está muy bien, pero si queremos utilizar el resultado en otra operación, puede ser un engorro, porque la respuesta no es un número. Afortunadamente el remedio es muy fácil. Sólo hay que añadir una pequeña modificación al final de la llamada a la función `integrate`:

```
integrate( f, lower=1/2, upper=3/4)$value
## [1] 0.3438
```

y ahora se obtiene como salida un número, que podemos asignar a una variable para usarlo en otras operaciones.

En ese primer ejemplo, el intervalo de integración $(\frac{1}{2}, \frac{3}{4})$ estaba completamente contenido dentro del soporte de la función f . Pero no hay ningún inconveniente en integrar sobre intervalos más grandes, que incluyan regiones externas al soporte. Por ejemplo, podemos integrar f en el intervalo $(\frac{1}{2}, 5)$, o incluso comprobar que es realmente una función de densidad, integrando en $(-\infty, \infty)$ (recuerda que en R se usa `Inf`):

```
integrate( f, lower=1/2, upper=5)$value
## [1] 0.5

integrate( f, lower=-Inf, upper=Inf)$value
## [1] 1
```

6.3.1. Cálculo de la media y varianza.

¿Y para calcular la media μ de la variable X cuya densidad es $f(x)$? Recordemos que, si la densidad de X en (a, b) es la función $f(x)$, entonces la media de X es:

$$\mu_X = \int_a^b x f(x) dx$$

Para aplicar esto a la función $f(x)$, como primer paso, creamos una función auxiliar que llamaremos `aux_f` (elegimos ese nombre para recordar su papel; podría ser cualquier otro), cuyo valor es el de $f(x)$ multiplicado por x .

```
aux_f = function(x){ x * f(x) }
```

Es interesante observar que la definición de esta función incluye una llamada a la propia función $f(x)$, en lugar de copiar directamente su definición. De esa manera, esta función auxiliar sigue siendo válida incluso aunque la función $f(x)$ original cambie.

Y ahora, para calcular la media μ de $f(x)$ basta con calcular la integral de esta función auxiliar entre $-\infty$ e ∞ .

```
(mu=integrate(aux_f, lower=-Inf, upper=Inf)$value)
## [1] 0.5
```

El siguiente paso lógico es obtener la varianza σ^2 (y desviación típica σ) de X . Ahora, recordando que la fórmula para la varianza es

$$\sigma_X^2 = \int_a^b (x - \mu)^2 f(x) dx$$

ya debería estar claro que el primer paso es definir una nueva función auxiliar:

```
aux2_f = function(x){ (x-mu)^2 * f(x) }
```

e integrarla entre $-\infty$ e ∞ .

```
(varianza=integrate(aux2_f, lower=-Inf, upper=Inf)$value )
## [1] 0.05
```

La desviación típica es simplemente la raíz cuadrada de este número (calculada, por ejemplo, con la función `sqrt` de R). Puedes comprobar que los resultados coinciden con los del apartado 2 del Ejercicio 15 (solución en la página 66), donde hicimos estas mismas integrales usando otro programa.

Ejercicio 28.

Usa R para comprobar los resultados del Ejemplo 5.4.6 del libro (pág. 162). Ver también el comienzo de la Sección 3.5 de este tutorial. Solución en la página 80. \square

6.3.2. Peligros de la integración numérica.

En la página 27 hemos discutido la no existencia de la media cuando la función de densidad es la función que en el Ejercicio 25 (pág. 44) hemos llamado `f2`. Queremos invitarte ahora a que compruebes lo que sucede si tratas de calcular esa media con R. El código sería este:

```
f2 = function(x){ 1 / (pi * (1 + x^2))}
aux_f2 = function(x){ x * f2(x) }
(mu=integrate(aux_f2, lower=-Inf, upper=Inf)$value)
## [1] 0
```

Y, como ves, la respuesta de R parece indicar que la media es 0. Pero ya vimos que en realidad la media no existe porque la integral que la define produce una indeterminación de la forma $-\infty - \infty$. De hecho, si le pedimos a R que calcule sólo la mitad derecha de la integral (de 0 a ∞), las dificultades se hacen evidentes:

```
integrate(aux_f2, lower=0, upper=Inf)$value
## Error in integrate(aux_f2, lower = 0, upper = Inf): maximum number of subdivisions
reached
```

Ahora R nos avisa de que no ha sido capaz de establecer el valor de esa integral. El problema, aquí, es que la simetría de la función que integramos está confundiendo a R, haciendo que la parte negativa se compense exactamente con la parte positiva. Eso es lo que hace que R piense que el resultado de la integral completa, de $-\infty$ a ∞ es 0. Y si nos hubiéramos limitado a usar R para esto, obtendríamos un resultado incorrecto, que podría conducirnos a errores serios más adelante.

“¿Y cómo puedo saberlo, cómo puedo saber yo que el programa ha fallado?”, te estarás preguntando. La respuesta, me temo, es que no puedes. Los programas mejoran cada día. Pero todavía es necesario

muchas veces recurrir a la ayuda de un experto para asegurarnos de que no hemos cometido un error por el camino. En cualquier caso, no te preocupes excesivamente. Este tipo de problemas tienen un impacto muy pequeño sobre la práctica diaria de los usuarios de la Estadística. Cuando tengas más experiencia, aprenderás a juzgar si ha llegado el momento de pedir ayuda a un estadístico (lo cual siempre es una buena idea, en proyectos que involucren un componente importante de análisis de datos).

6.4. Función de distribución (acumulada) para una variable aleatoria continua.

La función de distribución (acumulada) de una variable aleatoria continua X , definida en el intervalo (a, b) por la función de densidad $f(x)$ es

$$F(x) = P(X \leq x) = \int_a^x f(t)dt$$

Para aclararlo un poco: si, por ejemplo, $(a, b) = (1, 5)$, y queremos calcular $F(3)$, entonces tenemos que integrar la densidad $f(x)$ desde 1 (el principio del intervalo) hasta 3 (el valor en el que calculamos F).

Esta función de distribución F se puede definir en R de forma sencilla, recurriendo al comando `integrate`. En lugar de llamarla `F`, que ya se usa como abreviatura para el valor lógico `FALSE` en R, y podría generar ambigüedades, vamos a llamarla `Df` en R. Si usamos la misma función de densidad `f` que venimos usando desde la página 43, tenemos:

```
Df = function(x){integrate(f, lower=-Inf, upper=x)$value}
```

Obsérva que de nuevo hemos usado `$value`, para obtener el valor de la integral. Ahora podemos pedirle a R que calcule cualquier valor de la función de distribución (acumulada) `Df`. Por ejemplo, se obtiene:

```
Df(1/3)
## [1] 0.2593
```

lo cual significa que:

$$P(X \leq 1/3) \approx 0.2593$$

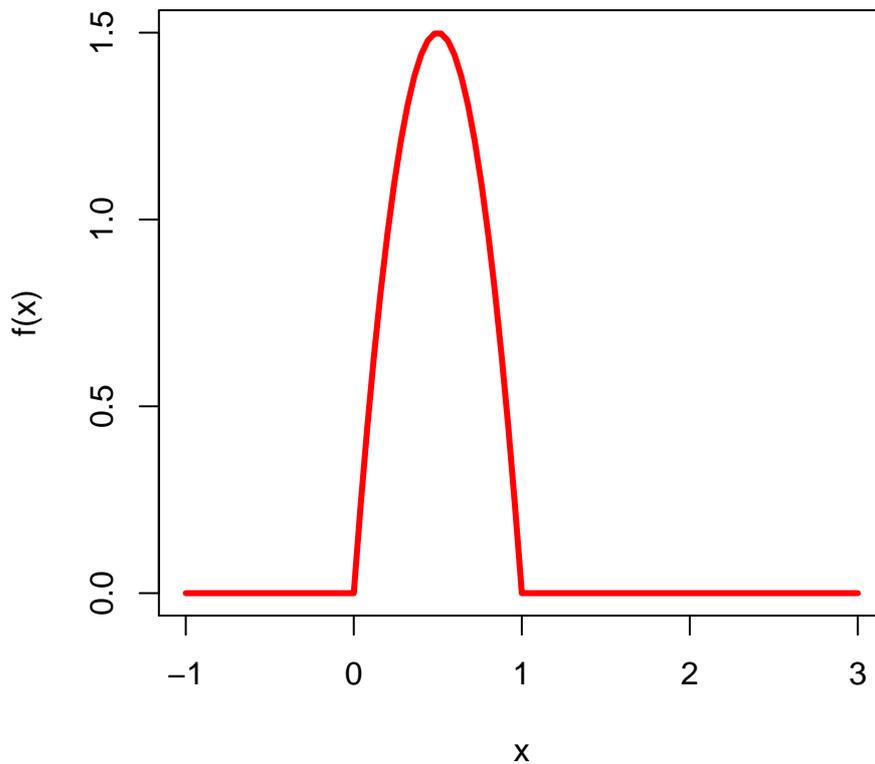
Recordemos que en el caso de la distribución normal, la función de distribución acumulada (con el mismo sentido que estamos usando aquí) se obtiene mediante `pnorm`. Y en general, para cualquier distribución *con nombre propio*, el prefijo `p` indica que calculamos la función de distribución. Lo que hemos aprendido aquí es cómo calcular la función de distribución acumulada para nuestras propias variables aleatorias continuas.

6.5. Gráficas de funciones con R.

No querríamos despedirnos de este estudio de las funciones con R sin discutir, al menos desde un punto de vista muy básico, cómo usar R para dibujar las gráficas de estas funciones.

Una forma muy sencilla de dibujar la gráfica de una función es usando la función `curve` de R. Por ejemplo, para la función `f` haríamos:

```
curve(f, from=-1, to=3, ylim=range(0, 1.5), col="red", lwd=3)
```



Compara el resultado con la primera figura de la Sección 3.4, obtenida con GeoGebra. Para poder usar `curve` es necesario que la función que vamos a representar se pueda aplicar a vectores (recuerda la discusión en torno a la función `ifelse` de la Sección 6.2, pág. 45). En futuros tutoriales aprenderemos a usar la función `plot`, que nos dará un control más fino sobre el resultado gráfico.

Ejercicio 29.

1. Dibuja, usando `curve`, la gráfica de la función `f2` del Ejercicio 25 (pág. 44).
2. Este apartado es más difícil, para que puedas practicar la definición de funciones a trozos en R , usando `ifelse`. Dibuja, usando `curve` la gráfica de la función de densidad del Ejemplo 5.5.2 del libro (pág. 168).

Soluciones en la página 81.

□

7. Ejercicios adicionales y soluciones.

Ejercicios adicionales

Distribución binomial.

Ejercicio 30. Un laboratorio farmacéutico está haciendo pruebas para evaluar la posible toxicidad de un nuevo tratamiento. Para ello se inyecta el producto a ratas. Se ha observado que 4 de cada 20 ratas mueren antes de que el experimento haya concluido. Teniendo esto en cuenta, si se inyecta el producto a 10 ratas, ¿cuál es la probabilidad de que al menos 8 de ellas sobrevivan? □

Ejercicio 31. Las encuestas electorales aseguran que, en las próximas elecciones, el 25% de los electores votarán al Partido de la Unidad y Felicidad Óptimas (PUFO). Se eligen al azar 10 electores. Calcular la probabilidad de que al menos 3 de ellos sean votantes del PUFO, y de que lo

sean al menos 9 de los 10. Calcula el valor esperado y la desviación típica de la variable $X = \{\text{número de votantes del PUFO entre los 10 elegidos}\}$.

□

Ejercicio 32. En cierta población de bacterias se ha comprobado que un el 0.06 % son, de hecho, superbacterias (poseen resistencia a los antibióticos). En una muestra de 10000 bacterias de dicha población, calcular:

1. La probabilidad de que haya exactamente 15 superbacterias.
2. La probabilidad de que el número de superbacterias sea superior a 10 pero inferior a 15.

□

Ejercicio 33. En un lote de 1000 piezas hay un 7% de piezas defectuosas. Se elige una muestra aleatoria de 50 piezas (con reemplazamiento). Calcula:

1. La probabilidad de que haya exactamente 8 piezas defectuosas.
2. La probabilidad de que haya a lo sumo 8 piezas defectuosas.
3. ¿Y si el muestreo se hubiera hecho sin reemplazamiento, cuáles serían esas probabilidades?

□

Ejercicio 34. Sea X una variable aleatoria con distribución $B(n, p)$, donde $n > 1$ es un número fijo, pero p es desconocido. ¿Cuál es el valor de p para el que la varianza de X es máxima? ¿Y cuál es ese valor máximo de la varianza?

□

Ejercicio 35. Si la probabilidad de acertar en un blanco es $1/5$, y se hacen 10 disparos de forma independiente. ¿Cuál es la probabilidad de acertar por lo menos dos veces, sabiendo que se acertó al menos una vez?

□

VARIABLES ALEATORIAS CONTINUAS.

Ejercicio 36. A continuación aparecen una serie de funciones. En todos los casos, se trata de:

- Dibujar la gráfica de la función (usa GeoGebra, Wolfram Alpha o algo similar).
- Encontrar, si existe, un valor de k para el que $f(x)$ sea una función de densidad.
- Si existe k , sea X la variable aleatoria continua definida por f para ese valor de k .
- Calcular las probabilidades que se indican en cada apartado. Recomendamos usar Wiris o Wolfram Alpha para calcular las integrales (a veces es preferible uno frente al otro; haz experimentos...).

Las funciones son estas:

1. $f(x) = k \cdot e^{-|x|}$.
Calcular las probabilidades $P(0 < X < 1)$, $P(-1 < X < 1)$, $P(X > 1)$.
2. $f(x) = \frac{k}{1 + |x|}$.
Calcular las probabilidades $P(0 < X < 1)$, $P(-1 < X < 1)$, $P(X > 1)$.
3. $f(x) = k \cdot \frac{2 + \sin(3x)}{1 + x^2}$.
Calcular las probabilidades $P(-3 < X < 1)$, $P(X > 0)$.
4. $f(x) = k \cdot \frac{x^2 - 2x + 2}{2x^4 + 5x^2 + 2}$.
Calcular las probabilidades $P(-1 < X < 2)$, $P(X > 3)$.

□

Ejercicio 37. Sea X una variable aleatoria continua que tiene esta función de densidad:

$$f(x) = \begin{cases} 1-x & \text{si } 0 \leq x < 1 \\ x-1 & \text{si } 1 \leq x \leq 2 \\ 0 & \text{en otro caso.} \end{cases}$$

Sean además los sucesos:

$$A = \{0 \leq X \leq 1\}.$$

$$B = \{-2 \leq X \leq 2\}.$$

$$C = \{\frac{1}{2} \leq X \leq +\infty\}.$$

$$D = \{X \text{ toma uno de estos valores: } 0, 1/2, 1, 2\}.$$

$$E = \{0.5 \leq X \leq 1.5\}$$

Calcular la probabilidad de los sucesos $A, B, C, D, E, A \cap C$. Calcular también $P(A|E)$.

□

Ejercicio 38. Sea X una variable aleatoria continua cuya función de densidad es:

$$f(x) = \begin{cases} k(1+x^2) & \text{si } 0 < x < 3 \\ 0 & \text{en otro caso.} \end{cases}$$

Calcular:

1. la constante k .
2. $P(1 < X < 2)$.
3. $P(X < 1)$.
4. $P(X > 1 | X < 2)$.

□

Ejercicio 39. Sea X una variable aleatoria uniforme (ver Capítulo 5, sección 5.4.3) en el intervalo $(1, 5)$. Calcula un valor a tal que

$$P(1+a < X < 5-a) = 0.9$$

□

Ejercicio 40. Una variable aleatoria X es exponencial si su función de densidad es de la forma:

$$f(x) = \begin{cases} 0 & \text{si } x < 0 \\ ke^{-kx} & \text{si } 0 \leq x \end{cases}$$

para alguna constante $k > 0$.

1. Comprueba que, sea cual sea el valor de k , la función f es, en efecto, una función de densidad.
2. Sea $k = 2$. Calcula la probabilidad $P(X > 1)$
3. Sea $k > 0$ un número cualquiera. Calcula $P(0 < X < k)$.

□

Ejercicio 41. Sea X una variable aleatoria de tipo uniforme (ver Capítulo 5, sección 5.4.3) en el intervalo (a, b) . Calcular la media de X , y comprobar que coincide con lo que predice la intuición. Calcular la varianza y desviación típica de X .

□

Ejercicio 42. Sea X una variable aleatoria de tipo exponencial (ver ejercicio 40). Calcular la media, varianza y desviación típica de X . □

Ejercicio 43. Calcula la media de las variables de los apartados (a) y (d) del Ejercicio 36. Calcula la varianza de la variable que aparece en el apartado (a) de ese ejercicio. □

Ejercicio 44. Sea X una variable aleatoria de tipo uniforme en el intervalo (a, b) . Calcular la función de distribución de X . □

Ejercicio 45. Sea X una variable aleatoria de tipo exponencial (ver ejercicio 40) Calcular la función de distribución de X . Úsala para rehacer el apartado (c) del Ejercicio 40. □

Ejercicio 46. Sea X una variable aleatoria cuya función de distribución es

$$F(x) = \frac{1}{1 + e^{-x}}.$$

Dibuja la gráfica de F (usa el ordenador). Párate un momento a pensar: ¿qué requisitos tiene que cumplir F para ser una función de distribución? Verifica que esta función F los cumple. Calcula las probabilidades $P(X < 3)$, $P(X > 2)$, $P(1 < X < 4)$. □

Distribución normal.

Ejercicio 47. La temperatura corporal en los adultos sanos sigue una distribución normal con media igual a 36.8 grados centígrados, y una desviación típica de 0.4 grados. Si elegimos un adulto sano al azar, calcular la probabilidad de que ocurra alguna de estas situaciones:

1. que su temperatura corporal sea mayor que 37 grados.
 2. que sea inferior a 35.5 grados.
 3. que se sitúe entre la media y 38 grados
 4. que esté comprendida entre 36 y 37 grados.
-

Ejercicio 48. Dada una distribución normal de media 3 y varianza 9, calcule las siguientes probabilidades:

1. $P(2 \leq X \leq 5)$.
 2. $P(X \geq 3)$.
 3. $P(X \leq -2)$.
-

Ejercicio 49. Calcule en los siguientes casos el valor de a , sabiendo que X es del tipo $N(1, 5)$.

1. $P(0 \leq X \leq a) = 0.28$
2. $P(1 - a < X < 1 + a) = 0.65$

Otros ejercicios.

□

Ejercicio 50. Las integrales aparecen en este curso porque las necesitamos para entender cómo funciona una variable aleatoria continua. Pero no son un objetivo central del curso y no nos vamos a demorar en ellas. Dicho esto, no nos podemos resistir a la tentación de usar las integrales para calcular el área de un círculo de radio r . En la escuela nos enseñan a calcular el área de un rectángulo, de un triángulo y en general de polígonos. Pero todas esas figuras se pueden descomponer en triángulos, así que en el fondo la única fórmula importante es la del área de un triángulo. En algún momento nos dicen también que el área de un círculo de radio r es $\pi \cdot r^2$. Pero el círculo no se puede descomponer en triángulos (al menos, en una cantidad finita de ellos), así que esa fórmula queda sin justificar para la mayoría de la gente. Vamos a usar las integrales para confirmarla.

Los puntos (x, y) de una circunferencia de radio r cumplen todos esta ecuación:

$$x^2 + y^2 = r^2.$$

Podemos despejar la y de esta ecuación así:

$$y = \pm\sqrt{r^2 - x^2}.$$

El signo \pm nos recuerda que para cada valor de x hay dos valores de la y : uno en la parte superior de la circunferencia (se obtiene con la raíz positiva) y otro en la parte inferior (con la raíz negativa). Eso es cierto salvo para $x = r$ o $x = -r$, porque en ese caso sólo hay un valor de la y (esto debería ser geoméricamente fácil de entender).

En cualquier caso, podemos usar estas ideas para escribir la ecuación de “la parte de arriba de la circunferencia” en forma de función:

$$y = f(x) = \sqrt{r^2 - x^2}$$

y usar esta función para calcular el área del semicírculo superior. Para hacerlo tenemos que calcular la integral:

$$\int_{-r}^r \sqrt{r^2 - x^2} dx$$

Tu trabajo en este ejercicio consiste en entender lo anterior y usar alguno de los programas que conoces para obtener el resultado y ver que es lo que esperábamos (recuerda que esa integral es el área de la mitad del círculo).

□

Ejercicio 51. Vamos a elegir dos números, X e Y , por orden. El número X , que se elige primero, sigue una distribución binomial $B(2, 1/3)$. Una vez elegido X , el número Y sigue una distribución normal $N(X, 1)$ (es decir, la media es el número X elegido en el primer paso). ¿Cuál es la probabilidad de que Y sea menor o igual que 1?

□

Ejercicio 52. En una fábrica hay 3 máquinas. La máquina $M1$ produce el 50 % del total de las piezas, la máquina $M2$ el 30 % y la máquina $M3$ el 20 % restante. Además, una pieza se considera defectuosa si su peso es mayor de 150g. El peso de las piezas fabricadas en $M1$ sigue una distribución normal de tipo $N(130, 20)$, el peso de las fabricadas en $M2$ sigue una $N(140, 15)$ y el de las fabricadas en $M3$ sigue una $N(140, 20)$. Si una pieza de esa fábrica resulta defectuosa, ¿cuál es la probabilidad de que proceda de $M2$?

□

Ejercicio 53. Vamos a jugar a un juego. Utilizamos una variable X que sigue la distribución normal $N(0, 1)$, y obtenemos un valor de X . Si se cumple $X \leq 0$, yo gano 2 euros. Si se cumple $0 < X \leq 0.6744898$, te pago un euro. Todavía hay que decidir lo que vamos a hacer cuando $X > 0.6744898$. Si queremos que el juego sea justo para ambos jugadores, ¿qué haremos en ese caso? Indicación: recuerda que un juego es justo si la ganancia media de los jugadores es 0.

□

Ejercicio 54. En una granja avícola han observado que el peso (en gramos) de los pollos de cuatro semanas sigue una distribución normal de tipo $N(\mu, \sigma)$ con $\mu = 1030, \sigma = 50$. La inspección sanitaria considera que los pollos cuyo peso es inferior a $\mu - 1.5 \cdot \sigma$ son no aptos, y deben ser apartados para recibir un tratamiento especial. Esta mañana, en una inspección de sanidad rutinaria, hemos elegido 100 pollos de cuatro semanas de esa granja (elección con reemplazamiento; una vez pesado el pollo se devuelve al corral y podría volver a ser elegido posteriormente). ¿Cuál es la probabilidad de que de esos 100 pollos haya al menos 10 no aptos?

□

Ejercicio 55.

En este ejercicio se trata de usar *R* para simular el proceso que se describe en el Ejemplo 5.1.3 del libro (pág. 135). Concretamente se trata de:

1. Generar un vector `resultados` con n simulaciones del proceso, donde n es un número muy grande (decenas o centenares de miles).
2. En cada iteración (usa un bucle `for`), empezamos por decidir (usa `ifelse`) qué urna usamos. La urna puede ser blanca "b", o negra "n".
3. Una vez decidida la urna, extremos una bola (número del 1 al 6) al azar, teniendo en cuenta la composición de la urna (usa `sample`).
4. Añadimos el valor de la bola al vector `resultados`.
5. Y según el valor de la bola, decidimos qué urna usaremos en la siguiente iteración.
6. Tras repetir n veces estos pasos, haz un diagrama de barras para la tabla de frecuencias del vector `resultados`.

Y a la vista de ese diagrama de barras, piensa si este proceso corresponde a una distribución binomial. Solución en la página 82.

□

Ejercicio 56.

En el Ejercicio 15 de este Tutorial (27) hemos visto que la distribución de Cauchy no tiene media. Es posible que ese resultado te haya intrigado. Si es así, este ejercicio es para ti, porque vamos a explorar esa idea con un poco más de detalle, usando *R*.

R incluye cuatro funciones para trabajar con la distribución de Cauchy que, de forma poco sorprendente, se llaman

`dcauchy`, `pcauchy`, `qcauchy`, `rcauchy`.

La primera (la función de densidad) apenas vamos a usarla. De hecho, la que nos interesa para este ejercicio es `rcauchy`. Vamos a usarla para explorar el comportamiento de las muestras aleatorias de la distribución de Cauchy.

Para entender el fenómeno que te vamos a mostrar, es bueno empezar con una distribución con comportamiento mucho más simple, como es la normal estándar $Z = N(0, 1)$. Si construimos una primera muestra aleatoria de valores de Z y calculamos la media de esos valores, obtendremos un valor muy cercano a 0:

```
set.seed(2014)
muestra1 = rnorm(10000)
mean(muestra1)

## [1] -0.003133
```

Imagínate que ahora calculamos otra muestra distinta de Z del mismo tamaño, pero le sumamos 5 a cada valor de la muestra.

```
muestra2 = 5 + rnorm(10000)
mean(muestra2)

## [1] 4.988
```

En cualquier caso, la diferencia entre las dos medias será muy aproximadamente igual a 5:

```
mean(muestra2) - mean(muestra1)

## [1] 4.991
```

Todo esto es casi tediosamente previsible. Para divertirnos un poco, lo que queremos que hagas en este ejercicio es repetir esta prueba, pero cambiando Z por la distribución de Cauchy. \square

Independencia y vectores aleatorios continuos.

Sólo debes intentar estos ejercicios si has leído la Sección 5.7 del libro.

Ejercicio 57.

El objetivo de este ejercicio es que uses el ordenador para comprobar los cálculos que aparecen en los ejemplos de esa sección. Personalmente, por comodidad y rapidez, te recomiendo que uses Wolfram Alpha. Para que te sirva de guía, la siguiente integral doble,

$$\int_{x=-\infty}^{\infty} \left(\int_{y=-\infty}^{\infty} \frac{1}{\pi} e^{-(x^2+y^2)} dy \right) dx = 1.$$

que aparece en el Ejemplo 5.7.1 (pág. 184) se puede calcular en Wolfram Alpha con este comando:

```
integrate (1/pi) * exp(-(x^2+y^2)) dx dy, y=-oo to oo, x=-oo to oo
```

Fíjate en que Wolfram Alpha usa `oo` (dos letras o minúsculas seguidas) como abreviatura de ∞ . Usando comandos parecidos, calcula $P(A)$ del Ejemplo 5.7.1 (pág. 184) y $f_X(x)$ del Ejemplo 5.7.3 (pág. 188). Solución en la página 84. \square

Ejercicio 58.

En este ejercicio sólo queremos que veas que R ofrece la posibilidad, mediante librerías adicionales, de visualizar gráficas tridimensionales de forma bastante satisfactoria. A la espera de lo que pueda ofrecer en ese sentido la próxima versión de GeoGebra, vamos a usar R para mostrarte una gráfica similar a la de la Figura 5.29 del libro (pág. 187), que además podrás rotar y cambiar de tamaño usando los botones del ratón.

Asegúrate de instalar las librerías `rgl` y `mnormt` de R antes de ejecutar el siguiente código (recuerda que vimos cómo instalar librerías en la Sección 6 del Tutorial03). Cuando lo ejecutes, el gráfico aparecerá en una ventana adicional. Te aconsejo que la maximices para poder explorar la figura con comodidad.

```
rm(list=ls())
library(rgl)
library(mnormt)

minx = -4
maxx = 4
miny = -4
maxy = 4

nx = 45
ny = 200

x= seq(minx, maxx, length.out = nx)
y= seq(miny, maxy, length.out = ny)

xnet = rep(x, ny)
ynet = rep(y, rep(nx,ny))

XY = matrix(c(xnet, ynet), ncol = 2)

znet = dmnorm(XY, varcov = matrix(c(1,0,0,1), ncol = 2))
```

```
plot3d(xnet, ynet, znet, col="blue")
points3d(ynet, xnet, znet, col="darkgreen")
```

□

Soluciones de algunos ejercicios

• Ejercicio 1, pág. 3

1. Para la primera parte, hacemos:

```
dbinom(0:10, size= 10, prob=1/5)

## [1] 1.074e-01 2.684e-01 3.020e-01 2.013e-01 8.808e-02 2.642e-02 5.505e-03
## [8] 7.864e-04 7.373e-05 4.096e-06 1.024e-07

max(dbinom(0:10, size= 10, prob=1/5))

## [1] 0.302
```

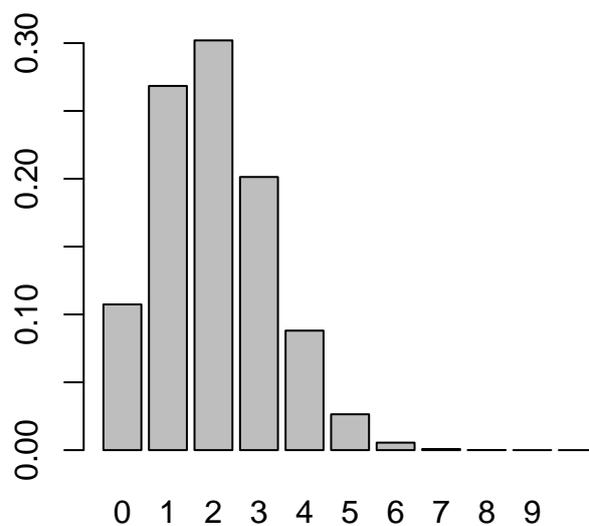
Para saber cual es el valor de k para el que se alcanza el máximo usamos una variante de la función `which`, que se llama `which.max` (también existe `which.min`, claro):

```
which.max(dbinom(0:10, size= 10, prob=1/5))

## [1] 3
```

Y puede ser una buena idea representar los valores gráficamente:

```
barplot(dbinom(0:10, size= 10, prob=1/5), names.arg = 0:10)
```



2. Lanzamos un dado cinco veces y el éxito se define como “sacar un seis”. Así que estamos trabajando con una binomial $X \sim B(n, p)$ en la que $n = 5$ y $p = \frac{1}{6}$. La probabilidad que queremos calcular en este ejemplo es:

$$P(X = 2)$$

que en R se obtiene mediante:

```
## [1] 0.1608
```

3. La tercera parte consiste en calcular $P(X = 3)$ en una binomial $B(11, \frac{1}{17})$:

```
dbinom(3, size=11, prob=1/17)
```

```
## [1] 0.02068
```

4. Sumamos los valores calculados con dbinom:

```
sum(dbinom(5:9, size=21, prob=1/3))
```

```
## [1] 0.7541
```

Fíjate en que aplicamos la función dbinom directamente al vector de valores 5:9 que nos interesan.

5.
6. Y para la última parte hacemos:

```
sum(dbinom(0:3, size=3, prob=1/5))
```

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

• Ejercicio 2, pág. 4

Fijamos los valores de n y p en este ejercicio:

```
n = 7  
p = 1/4
```

y empezamos con los cálculos.

1. Es directo:

```
pbinom(4, size = n, prob = p)
```

```
## [1] 0.9871
```

2. $P(X < 3) = P(X \leq 2)$, luego:

```
pbinom(2, size = n, prob = p)
```

```
## [1] 0.7564
```

3. $P(X \geq 2) = 1 - P(X < 2) = 1 - P(X \leq 1)$. ¡Cuidado en el segundo paso!

```
1- pbinom(1, size = n, prob = p)
## [1] 0.5551
```

De otra manera, usando `dbinom` y sumando:

```
sum(dbinom(2:n, size = n, prob = p))
## [1] 0.5551
```

4. $P(X > 3) = 1 - P(X \leq 3)$

```
1- pbinom(3, size = n, prob = p)
## [1] 0.07056
```

5. Podemos usar que $P(2 \leq X \leq 4) = P(X \leq 4) - P(x \leq 1)$

```
pbinom(4, size = n, prob = p) - pbinom(1, size = n, prob = p)
## [1] 0.5422
```

Compruébalo sumando valores de `dbinom`.

6. Ahora usamos $P(2 < X < 4) = P(X \leq 3) - P(x \leq 2) = P(X = 3)$

```
pbinom(3, size = n, prob = p) - pbinom(2, size = n, prob = p)
## [1] 0.173
dbinom(3, size = n, prob = p)
## [1] 0.173
```

El resultado es el mismo, claro.

7. $P(2 \leq X < 4) = P(X \leq 3) - P(x \leq 1)$

```
pbinom(3, size = n, prob = p) - pbinom(1, size = n, prob = p)
## [1] 0.4845
```

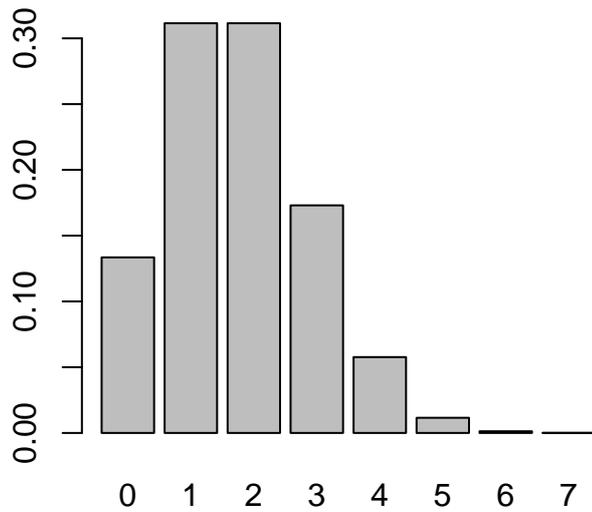
8. $P(2 < X \leq 4) = P(X \leq 4) - P(x \leq 2)$

```
pbinom(4, size = n, prob = p) - pbinom(2, size = n, prob = p)
## [1] 0.2307
```

• Ejercicio 3, pág. 5

Los valores y las correspondiente gráficas son, para `dbinom`:

```
dbinom(0:n, size=n, prob=p)
## [1] 0.13348389 0.31146240 0.31146240 0.17303467 0.05767822 0.01153564
## [7] 0.00128174 0.00006104
barplot(dbinom(0:n, size=n, prob=p), names.arg = 0:n)
```

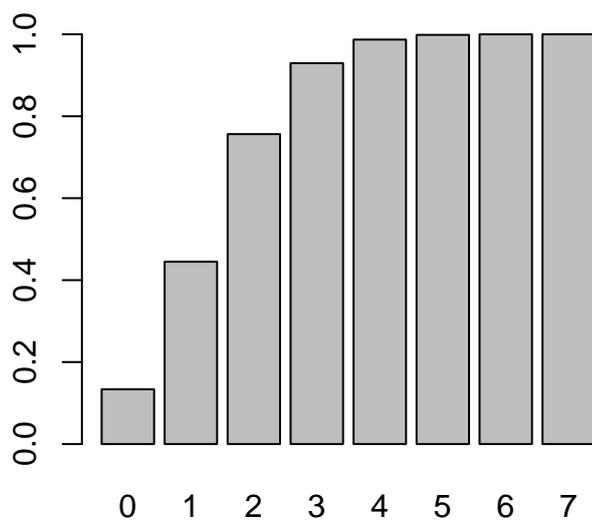


Y para pbinom

```
dbinom(0:n, size=n, prob=p)

## [1] 0.13348389 0.31146240 0.31146240 0.17303467 0.05767822 0.01153564
## [7] 0.00128174 0.00006104

barplot(pbinom(0:n, size=n, prob=p), names.arg = 0:n)
```



• **Ejercicio 4, pág. 5**

En general, dado cualquier valor de k entre 0 y n , se cumple que:

$$P(X \geq k) = 1 - P(X < k) = 1 - P(X \leq (k - 1)).$$

Por si te lo preguntas, para $k = 0$ es

$$P(X \geq 0) = 1 - P(X \leq -1) = 1 - 0,$$

que, si lo piensas, es lo que cabría esperar.

Así que podemos empezar calculando:

```
(probabilidades = 1- pbinom((0:7)-1, size = 7, prob = 1/4))  
  
## [1] 1.00000000 0.86651611 0.55505371 0.24359131 0.07055664 0.01287842  
## [7] 0.00134277 0.00006104
```

Y ahora que tenemos estos valores, sólo hay que compararlos con 0.75, localizar el mayor valor de k para el que se cumple la condición. Recuerda, de nuevo, que las posiciones en los vectores de R se cuentan desde 1, mientras que k empieza en 0, así que hay que restar 1 para obtener k .

```
max(which(probabilidades >= 0.75) )- 1  
  
## [1] 1
```

• **Ejercicio 5, pág. 6**

El primer paso sería:

```
(probabilidades = pbinom((0:7)-1, size = 7, prob = 1/4, lower.tail=FALSE))  
  
## [1] 1.00000000 0.86651611 0.55505371 0.24359131 0.07055664 0.01287842  
## [7] 0.00134277 0.00006104
```

Los resultado son los mismos que antes, así que a partir de aquí el resto del del ejercicio es igual. Al usar la cola derecha la expresión se simplifica un poco, como ves.

• **Ejercicio 6, pág. 6**

Vamos a fijar los valores de n y p para los distintos apartados:

```
n = 10  
p = 1/5
```

Y ahora vamos con las cuentas:

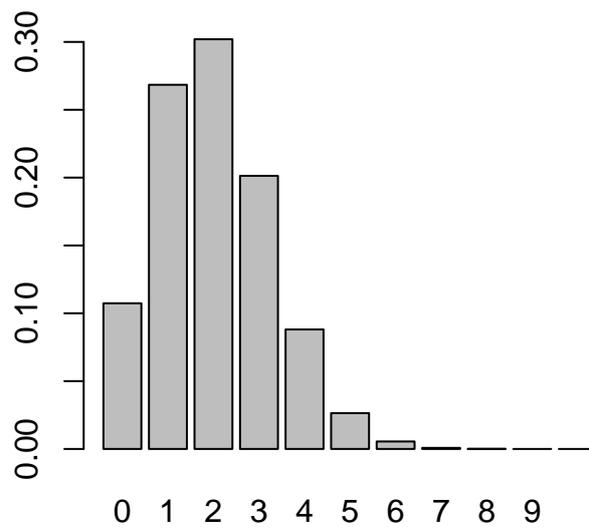
1. `1-pbinom(8, size=n, prob=p)`

```
## [1] 0.000004198
```

2. `dbinom(0:10, size=n, prob=p)`

```
## [1] 1.074e-01 2.684e-01 3.020e-01 2.013e-01 8.808e-02 2.642e-02 5.505e-03  
## [8] 7.864e-04 7.373e-05 4.096e-06 1.024e-07
```

```
barplot(dbinom(0:10, size=n, prob=p), names.arg = 0:10)
```



Como ves, la probabilidad se concentra en los primeros valores.

```
3. set.seed(2014)

numeros = rbinom(200, size=n, prob=p)
mean(numeros)

## [1] 2.125

sd(numeros)

## [1] 1.36

## Teoricos:
(mu = n * p)

## [1] 2

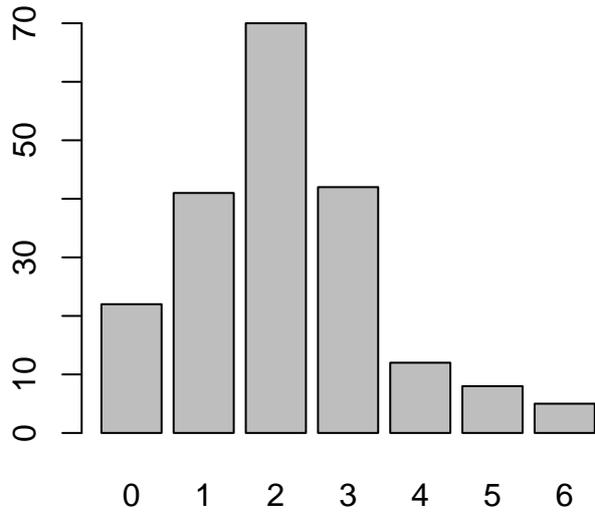
(sigma = sqrt (n * p * (1 - p)))

## [1] 1.265
```

```
4. table(numeros)

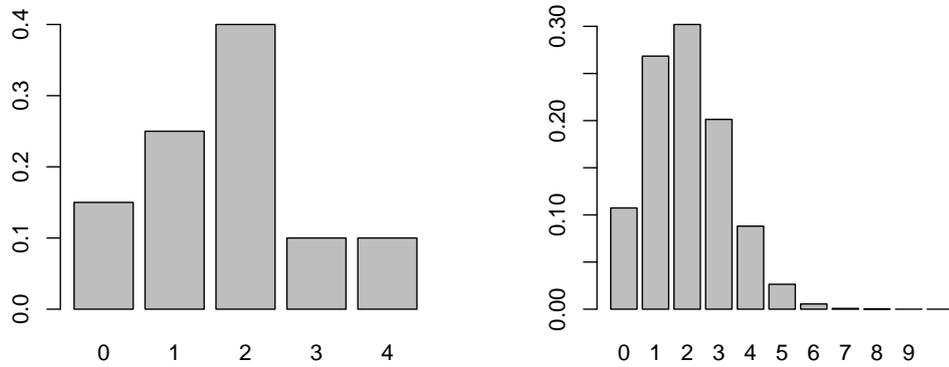
## numeros
## 0 1 2 3 4 5 6
## 22 41 70 42 12 8 5

barplot(table(numeros), names.arg = 0:max(numeros))
```

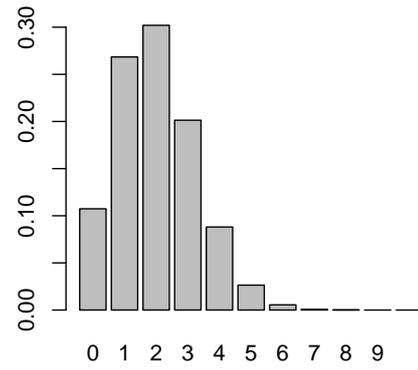
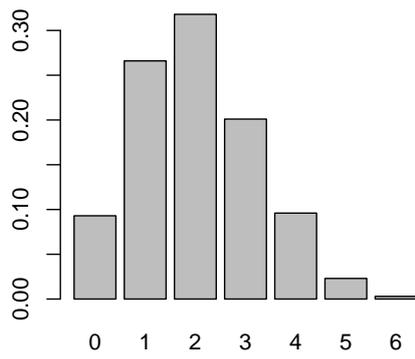


¿Por qué hemos usado `0:max(numeros)` en lugar de `0:n` al rotular el eje horizontal?

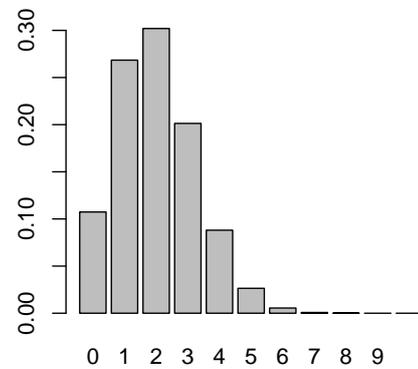
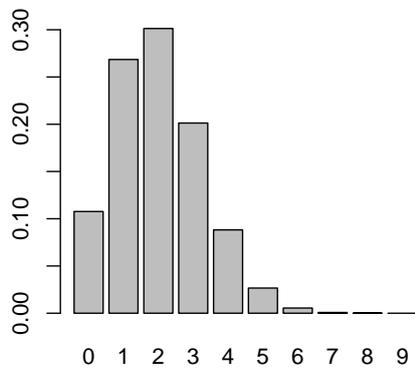
5. El resultado de ejecutar el código con 20 números aleatorios es :



A la derecha se muestra la teoría (la población, si prefieres pensarlo así), mientras que a la izquierda vemos la muestra. Repitiendo esto con 1000 valores en la muestra obtenemos:



Y si usamos $n = 1000000$:



Los dos gráficos son prácticamente idénticos.

• Ejercicio 7, pág. 7

1. Primer apartado:

```
n = 3
p = 1/5

probabilidades = dbinom(0:n, size=n, prob=p)
valores = 0:n

(mu = sum(valores * probabilidades))

## [1] 0.6

(sigma2 = sum((valores - mu)^2 * probabilidades))

## [1] 0.48

(sigma = sqrt(sigma2))

## [1] 0.6928
```

En fracciones:

```
library(MASS)
fractions(mu)

## [1] 3/5

fractions(sigma2)

## [1] 12/25
```

que coincide con lo que aparece en el ejemplo del libro.

En el panel de vista simbólica de GeoGebra estos resultados se obtienen con:

```
mu:= Suma[k * NúmeroCombinatorio[3, k]*(1/5)^k*(4/5)^(3-k), k, 0, 3]
```

y

```
sigma2 := Suma[(k - mu)^2 * NúmeroCombinatorio[3, k]*(1/5)^k*(4/5)^(3-k), k, 0, 3]
```

respectivamente.

2. Segundo apartado:

```
set.seed(2014)
# n es el tamaño de la muestra
n = 50000
# Los parámetros de la binomial son
N = 3
p = 1/5
# usamos rbinom
muestra = rbinom(n, size=N, prob = p)
#Y la media de la muestra es:
mean(muestra)

## [1] 0.5982
```

Como ves, hemos obtenido un valor próximo al valor teórico $0.6 \approx 0.6$. Dejamos al lector que haga las modificaciones necesarias de este código para llevar adelante el resto del experimento propuesto en el ejercicio.

• Ejercicio 9, pág. 8

En el panel de cálculo simbólico de GeoGebra puedes usar el comando:

```
Suma[NúmeroCombinatorio[1000, k]*(1/3)^k*(2/3)^(1000-k), k, 300, 600]
```

y usarla barra de deslizamiento para ver la fracción que se obtiene...

• Ejercicio 13, pág. 25

1. En Wiris se obtiene:

$$\left[\int_0^1 \frac{1}{\pi \cdot (1+x^2)} dx \rightarrow \frac{1}{4} \right]$$
$$\left[\int_1^{+\infty} \frac{1}{\pi \cdot (1+x^2)} dx \rightarrow \frac{1}{4} \right]$$

En Wolfram Alpha los comandos

```
integrate(1 / (pi * (1 + x^2))) from 0 to 1
```

y

```
integrate(1 / (pi * (1 + x^2))) from 1 to infinity
```

producen los resultados deseados.

2. El valor se calcula así:

$$\int_{\frac{1}{2}}^{\frac{3}{4}} 6 \cdot (x - x^2) dx \rightarrow \frac{11}{32}$$

3. Como muestra esta figura:

$$\int 6 \cdot (x - x^2) dx \rightarrow -2 \cdot x^3 + 3 \cdot x^2$$

la primitiva es $F(x) = -2x^3 + 3x^2$

4. Utiliza el siguiente comando en la *Línea de Entrada*:

```
Integral[6 * (x - x^2), 0, 1]
```

5. El resultado en Wiris es:

$$\int_0^1 6 \cdot (x - x^2) dx \rightarrow 1$$

En Wolfram Alpha tienes que usar el comando:

```
integrate(6 * (x - x^2) ) from 0 to 1
```

• **Ejercicio 14, pág. 26**

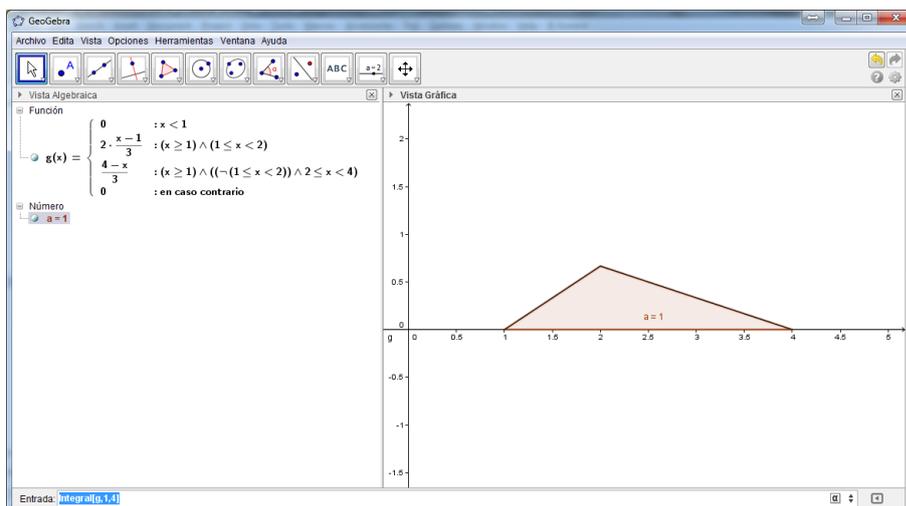
1. El resultado es aproximadamente 0.6064.
2. El resultado es aproximadamente 0.784.
3. Aunque no es la única manera, nosotros hemos usado este comando en GeoGebra para definir la función

$$g(x) = \text{Si}[x < 1, 0, \text{Si}[1 \leq x < 2, 2 \cdot (x-1)/3, \text{Si}[2 \leq x < 4, (4-x)/3, 0]]]$$

Para comprobar que g es una función de densidad basta con hacer:

```
Integral[g,1,4]
```

y se obtiene 1. En la siguiente figura se ilustran la definición de g y esa comprobación:



Verás que GeoGebra usa los operadores booleanos \wedge , que significa “y” (como el & de R), y \neg , que significa “no”, como el ! de R. Para calcular la probabilidad pedida ejecutamos el comando

$$\text{Integral}[g, 1879/1250, 2511/625]$$

y el resultado es aproximadamente 0.9156.

¿Crees que sabrías usar Wolfram Alpha para reproducir estos resultados? Una ventaja sería que la respuesta será simbólica, y por lo tanto exacta, no aproximada.

• **Ejercicio 15, pág. 27**

1. En Wolfram Alpha usamos los comandos:

$$\text{integrate}(x^2 * (2 - x)) \text{ from } 1 \text{ to } 2$$

$$\text{integrate}((x - 4/3)^2 * 2 * (2 - x)) \text{ from } 1 \text{ to } 2$$

para obtener, respectivamente, la media y la varianza. En Wiris es:

$$\left\{ \begin{array}{l} \mu = \int_1^2 x \cdot 2 \cdot (2-x) dx \rightarrow \frac{4}{3} \\ \sigma = \int_1^2 (x - \mu)^2 \cdot 2 \cdot (2-x) dx \rightarrow \frac{1}{18} \end{array} \right.$$

Aunque no era, desde luego, necesario, hemos preferido usar los símbolos μ y σ del menú *Griego* de Wiris.

2. Usando en Wolfram Alpha los comandos:

$$\text{integrate}(x^6 * (x - x^2)) \text{ from } 0 \text{ to } 1$$

$$\text{integrate}((x - 1/2)^2 * 6 * (x - x^2)) \text{ from } 0 \text{ to } 1$$

el primero nos muestra que la media es $\frac{1}{2}$, y entonces usamos ese resultado en el segundo para ver que la varianza es $\frac{1}{20} = 0.05$

3. En GeoGebra, si ejecutas el comando

$$\text{Integral}[x/(pi * (1 + x^2)), -\infty, \infty]$$

obtendrás como resultado **indefinido** (o incluso peor, la respuesta puede ser 0, dependiendo de la versión del programa que uses). Eso no aclara mucho las cosas, pero si pruebas con la mitad derecha de la integral, usando:

`Integral[x/(pi * (1 + x^2)), 0, ∞]`

obtendrás ∞ como respuesta. Ahora, la simetría de la gráfica de la función que integras hace evidente que el lado izquierdo de la integral (desde $-\infty$ hasta 0) vale $-\infty$. Por eso antes nos hemos tropezado con **indefinido**, porque GeoGebra se ha encontrado en una de esas situaciones del tipo $\infty - \infty$.

En Wolfram Alpha puedes usar el comando

`integrate(x/(pi * (1 + x^2))) from -infinity to infinity`

y la primera respuesta que obtendrás es *integral does not converge*, que es la forma en la que Wolfram Alpha nos informa de que algo ha ido mal con esta integral.

• **Ejercicio 16, pág. 30**

1. El resultado en cualquiera de los dos casos es:

$$1/2k^2 + c,$$

siendo c una constante. Es decir, que no es lo que queríamos obtener.

2. En el panel de *Cálculo Simbólico* de GeoGebra hemos ejecutado estos comandos:

`f(x) := x * cos(3*x)`
`Integral[f(x), x]`

y el resultado es la primitiva

$$F(x) = \frac{1}{9} \cos(3x) + \frac{1}{3}x \operatorname{sen}(3x) + c$$

siendo c una constante. Usando Wolfram Alpha, el comando:

`integrate(x * cos(3 * x))`

produce el mismo resultado.

3. Hay, básicamente, dos formas de abordar esta parte del ejercicio. En primer lugar, podemos usar la primitiva F que hemos obtenido en el anterior apartado y buscar el valor de k para el que se cumple (en el segundo paso, sacamos k de la integral):

$$1 = \int_0^{\frac{\pi}{6}} k \cdot x \cdot \cos(3x) dx = k \cdot \int_0^{\frac{\pi}{6}} x \cdot \cos(3x) dx = k \left(F\left(\frac{\pi}{6}\right) - F(0) \right).$$

Esto es:

$$1 = k \cdot \left(\frac{\pi}{18} - \frac{1}{9} \right),$$

de donde se deduce que $k = \frac{18}{\pi - 2} \approx 15.77$. En segundo lugar, tras sacar k de la integral, podemos despejar directamente:

$$k = \frac{1}{\int_0^{\frac{\pi}{6}} x \cdot \cos(3x) dx},$$

y usar cualquiera de los programas que hemos visto para calcular este valor de k . Por ejemplo, en Wolfram Alpha ejecutamos el comando:

`1 / integrate(x * cos(3 * x), x=0, x=pi/6)`

para obtener el mismo resultado.

4. Para hacer la integral del Ejemplo 5.5.1 en Wolfram Alpha puedes ejecutar el comando:

```
integrate( 1 / (pi * (1 + x^2))) from x= -infinity to x=k
```

En GeoGebra, prueba a ejecutar, en el panel de *Cálculo Simbólico*, estos comandos:

```
f(x) :=1/(pi*(1+x^2))
Integral[f(x), x, -∞, k]
```

La respuesta es la misma, escrita de forma ligeramente distinta:

$$\frac{\pi + 2 \arctan(k)}{2 \pi}$$

Para el Ejemplo 5.5.2, en Wolfram Alpha escribimos:

```
integrate 2*x/3 from 0 to k
```

Y se obtiene la respuesta esperada, $k^2/3$. La otra integral necesaria es

$$\int_2^k \frac{4}{3}(3-x)dx$$

que se obtiene en Wolfram Alpha con:

```
integrate (4/3)*(3 - x) from 2 to k
```

La respuesta es: $-\frac{2}{3} \cdot (k^2 - 6k + 8)$.

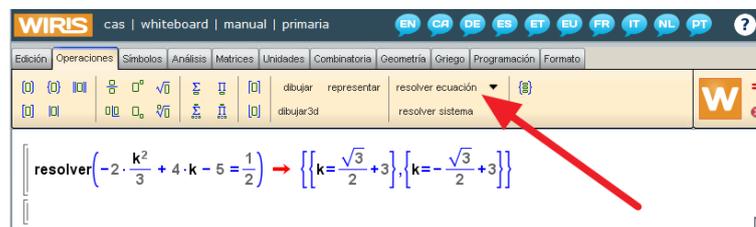
5. En Wolfram Alpha la ecuación se resuelve con:

```
solve( -2 * k^2/3 + 4 * k - 5 = 1/2)
```

En GeoGebra tienes que abrir el panel de *Cálculo Simbólico* y ejecutar este comando:

```
Resuelve[-2 * k^2/3 + 4 * k - 5 = 1/2]
```

Finalmente, en Wiris usamos el comando que aparece en la siguiente figura (hemos señalado dónde puedes localizar ese comando en la barra de menús)



• Ejercicio 17, pág. 31

- El resultado se obtiene con:

```
punif(6/130, min=1/130, max=12/130)
```

```
## [1] 0.4545
```

- En este caso hay que hacer una diferencia:

```
punif(9/130, min=1/130, max=12/130) - punif(2/130, min=1/130, max=12/130)
## [1] 0.6364
```

- Al tratarse de una variable aleatoria continua (la probabilidad de cada valor individual es 0), no hay ninguna diferencia.

```
dunif(6/130, min=1/130, max=12/130)
## [1] 11.82
dunif(20/130, min=1/130, max=12/130)
## [1] 0
```

Obviamente el primer valor no puede ser una probabilidad. Y el segundo es 0, porque estamos fuera del soporte.

- Basta con tener en cuenta que $P\left(X > \frac{6}{130}\right) = 1 - P\left(X \leq \frac{6}{130}\right)$.

```
1 - punif(6/130, min=1/130, max=12/130)
## [1] 0.5455
```

También se puede usar la opción `lower.tail=FALSE`

```
punif(6/130, min=1/130, max=12/130, lower.tail=FALSE)
## [1] 0.5455
```

• Ejercicio 18, pág. 32

```
(k = qunif(0.75, min=-11, max=25))
## [1] 16
punif(k, min=-11, max=25)
## [1] 0.75
```

• Ejercicio 19, pág. 32

Para resolverlo usamos que $P(X > k) = 1 - P(X \leq k)$, así que estamos buscando el valor de k tal que

$$1 - P(X \leq k) = \frac{1}{7},$$

o, lo que es lo mismo:

$$P(X \leq k) = 1 - \frac{1}{7}.$$

Ahora es fácil obtener la respuesta en R:

```
qunif((1 - (1/7)), min=-2, max=8)
```

```
## [1] 6.571
```

Otra manera de conseguirlo es con `lower.tail=FALSE`

```
qunif(1/7, min=-2, max=8, lower.tail=FALSE)
```

```
## [1] 6.571
```

Y para comprobar el resultado puedes hacer:

```
(k = qunif(1/7, min=-2, max=8, lower.tail=FALSE))
```

```
## [1] 6.571
```

```
punif(k, min=-2, max=8, lower.tail=FALSE)
```

```
## [1] 0.1429
```

```
1/7
```

```
## [1] 0.1429
```

• Ejercicio 20, pág. 32

1. Para generar ese vector hacemos:

```
set.seed(2014)
n = 1000
puntos = runif(n, min=-5, max=5)
```

Para hacernos una idea del resultado usamos `head` y `tail`:

```
head(puntos)
## [1] -2.1419 -3.3109 1.2591 -1.9031 0.4985 -4.1517

tail(puntos)
## [1] -0.3135 1.7720 1.5512 -2.2124 -0.5153 -3.5683
```

2. El cálculo con `punif` es este:

```
punif(3, min=-5, max=5) - punif(-1, min=-5, max=5)
## [1] 0.4
```

3. Empezamos construyendo un vector booleano (un vector de valores `TRUE/FALSE`) que nos indique si el punto ha caído en el intervalo `[-1, 3]` (recuerda que el operador booleano `&` se lee “y”):

```
enIntervalo = (puntos > -1 ) & (puntos < 3)
```

Y ahora usamos la equivalencia de TRUE/FALSE con 1/0 para sumar los valores TRUE:

```
(cuantosEnIntervalo = sum(enIntervalo))
```

```
## [1] 433
```

La fracción del total n de puntos que pertenecen al intervalo $[-1, 3]$ se obtiene con:

```
cuantosEnIntervalo / n
```

```
## [1] 0.433
```

4. Repetimos las operaciones con $n = 10000$, y le dejamos al lector la tarea de experimentar con valores mayores de n .

```
n = 10000
puntos = runif(n, min=-5, max=5)
enIntervalo = (puntos > -1 ) & (puntos < 3)
(cuantosEnIntervalo = sum(enIntervalo))
```

```
## [1] 3920
```

```
cuantosEnIntervalo / n
```

```
## [1] 0.392
```

También es una buena idea probar con un valor mucho más pequeño de n , como $n = 50$. ¿Qué sucede en ese caso?

5. El código de la simulación se muestra a continuación. Puede resultarte un poco más difícil que otros ejemplos anteriores, especialmente porque la parte gráfica usa muchas funciones de R que aún no hemos discutido. LA mejor forma de usar este código es copiándolo en el editor de texto de RStudio y ejecutando las instrucciones una a una, para ver las sucesivas figuras:

```
set.seed(2014)
```

```
# Empezamos dibujando un cuadrado de lado 2 y el círculo de radio 1  
# en su interior.
```

```
plot(c(seq(-1, 1, length.out = 1000), rep(1, 1000), seq(-1, 1, length.out = 1000), rep(-1,  
  , c(rep(-1, 1000), seq(-1, 1, length.out = 1000), rep(1, 1000), seq(-1, 1, length.out = 1000),  
  , xlab="", ylab="", bty="n")
```

```
curve(sqrt(1-x^2), from = -1, to = 1, lwd=3, col="blue", add=TRUE)  
curve(-sqrt(1-x^2), from = -1, to = 1, lwd=3, col="blue", add = TRUE)
```

```
# Vamos a lanzar n puntos al interior de un cuadrado de lado 2.  
n = 40
```

```
# En realidad, los vamos a elegir de entre los nodos de una malla rectangular  
# de n x n puntos superpuesta sobre el cuadrado.
```

```
# Construyo las coordenadas x de los puntos de la malla (cruces rojas en la figura).  
valoresx = seq(-1, 1, length.out = n + 1 )  
head(valoresx)
```

```
## [1] -1.00 -0.95 -0.90 -0.85 -0.80 -0.75
```

```

tail(valoresx)

## [1] 0.75 0.80 0.85 0.90 0.95 1.00

# Y sus coordenadas y
valoresy = seq(-1, 1, length.out = n + 1)

points(rep(valoresx,n+1),rep(valoresy, rep(n+1,n+1)), pch="+", cex=1.2, col="red")

# Los puntos donde caen los dardos se obtienen eligiendo su coordenadas con sample (cruces
x = sample(valoresx, n, replace = TRUE)
y = sample(valoresy, n, replace = TRUE)

points(x,y, cex=1.8, col="blue", pch="+")

# Y ahora me quedo sólo con los del círculo (cruces negras)
enCirculo = ( x^2 + y^2 < 1)

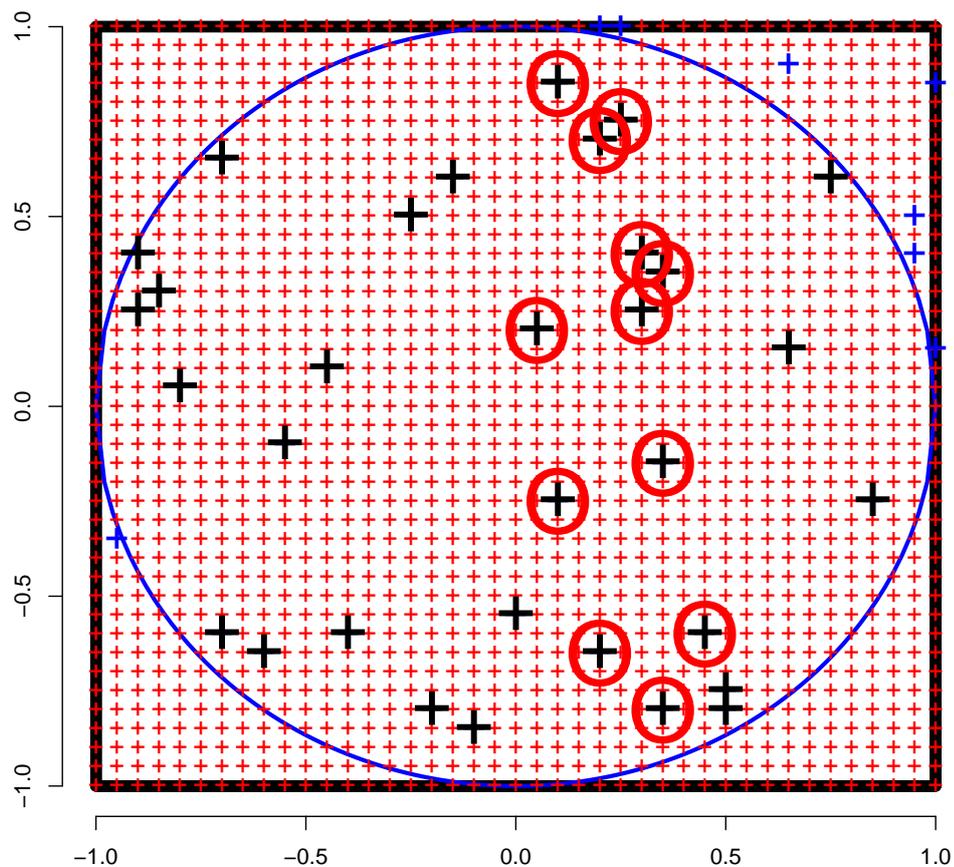
xC = x[enCirculo]
yC = y[enCirculo]
points(xC,yC, pch = "+", cex=3)

# El suceso A lo forman los puntos cuya coordenada X está ente 0 y 1/2.
A = (0 < xC) & (xC < 1/2)
xA =xC[A]
yA =yC[A]
head(xA)

## [1] 0.25 0.10 0.20 0.20 0.30 0.35

# Señalamos esos puntos rodeándolos con un círculo rojo.
points(xA, yA, col="red", pch="O", cex=4)

```



```
# Finalmente, la estimación de la probabilidad es:
(pA = sum(A) / sum(enCirculo))

## [1] 0.375

# Y el área de esa parte del círculo es:
pi * pA

## [1] 1.178
```

El código anterior está bien para entender cómo funciona la idea. Pero si vas a lanzar muchos dardos, las figuras empiezan a resultar muy enmarañadas. Si vas a lanzar unos miles de dardos, te conviene usar este otro fichero:

[Tut05-Ejemplo-4-1-2-Libro-SegundaParte.R](#)

Para finalizar el recorrido, hemos eliminado toda la parte gráfica de ese código y hemos “lanzado” nada menos que diez millones de dardos (nadie dijo que fuese a ser una tarea fácil calcular π por este método):

```
set.seed(2014)
n = 10000000
x = runif(n, min=-2, max=2)
y = runif(n, min=-2, max=2)
enCirculo = (x^2 + y^2 < 1)
(cuantosEnCirculo = sum(enCirculo))
```

```
## [1] 1963487

cuantosEnCirculo / n

## [1] 0.1963

16 * cuantosEnCirculo / n

## [1] 3.142
```

Como ves, el resultado es exacto hasta las cuatro primeras cifras significativas ($\pi \approx 3.1416$).

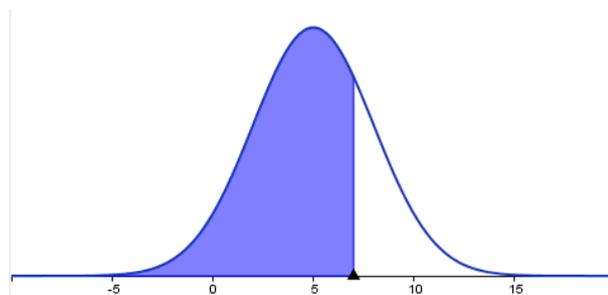
• **Ejercicio 21, pág. 35**

Todas las respuestas son aproximadas.

1. 0.3694, 0.2525 y 0.0478. A medida que nos desplazamos hacia la derecha los valores van siendo más y más pequeños.
2. 0.3694, 0.2525 y 0.0478. Las respuestas son las mismas del anterior apartado, por la simetría de la curva normal respecto de la media, que está en $\mu = 5$. Por ejemplo, los valores 3 y 7 están a la misma distancia, a izquierda y derecha de μ , respectivamente, y por eso

$$P(X < 3) = P(X > 7) = 0.2525$$

3. 0.1324, 0.6827 y 0.9044.
4. $k_1 = 8.8447$ y $k_2 = 9.9346$.
5. Los mismos valores de k_1 y k_2 , de nuevo por la simetría de la curva normal.
6. Es mayor que $1/2$. Se tiene $P(X < 7) = 0.7475$. Este apartado y el siguiente se resuelven teniendo en cuenta que el área de cada una de las dos mitades de la curva es $\frac{1}{2}$, observando si el punto que hemos tomado está a la derecha o la izquierda de μ , y si la probabilidad que calculamos incluye todos los valores mayores o menores. Por ejemplo, para la primera pregunta tenemos que pensar en un dibujo como este:



con el que resulta evidente que la respuesta es mayor que $\frac{1}{2}$.

7. Es menor que $1/2$. Se tiene $P(X > 8) = 0.1587$.
8. El valor $P(X > 4)$ es más grande. Se tiene $P(X > 4) = 0.6306$, mientras que $P(X < 5.5) = 0.5662$. En este caso la respuesta es fácil de ver porque el valor 4 está más lejos de $\mu = 5$ que 5.5.
9. Las ideas para este apartado y el siguiente son las mismas, pero ahora tenemos que pensar en los valores del eje x , en vez de pensar en las probabilidades (áreas) que definen esos valores. El valor tiene que ser menor que 5 (de otra manera, la probabilidad sería menor que $\frac{1}{2}$). Se obtiene $k = 4.24$
10. El valor tiene que ser menor que 5. Se obtiene $k = 1.156$.

• Ejercicio 22, pág. 39

1. Fijamos los valores de μ y σ para los tres apartados del Ejercicio 21.

```
mu = 5
sigma = 3
```

Y ahora vamos con los cálculos.

a) Primer apartado:

```
1 - pnorm(6, mean=mu, sd=sigma)
## [1] 0.3694

1 - pnorm(7, mean=mu, sd=sigma)
## [1] 0.2525

1 - pnorm(10, mean=mu, sd=sigma)
## [1] 0.04779
```

Fíjate en que podríamos haber calculado todos de una vez:

```
1 - pnorm(c(6, 7, 10), mean=mu, sd=sigma)
## [1] 0.36944 0.25249 0.04779
```

y que podríamos evitar restar de 1 con la opción `lower.tail`:

```
pnorm(c(6, 7, 10), mean=mu, sd=sigma, lower.tail=FALSE)
## [1] 0.36944 0.25249 0.04779
```

Compáralos con los resultados que obtuvimos con GeoGebra.

b) En una única instrucción:

```
pnorm(c(4, 3, 0), mean=mu, sd=sigma)
## [1] 0.36944 0.25249 0.04779
```

c) También es posible hacer los tres cálculos en una sola instrucción:

```
pnorm(c(5.5, 8, 10), mean=mu, sd=sigma) - pnorm(c(4.5, 2, 0), mean=mu, sd=sigma)
## [1] 0.1324 0.6827 0.9044
```

Fíjate en que el vector de la izquierda contiene los extremos superiores de los tres intervalos, mientras que el de la derecha contiene los tres extremos inferiores.

2. En este caso tenemos:

```
mu = 7
sigma = sqrt(14/3)
pnorm(9, mean=mu, sd=sigma) - pnorm(5, mean=mu, sd=sigma)
## [1] 0.6455

pnorm(9.5, mean=mu, sd=sigma) - pnorm(4.5, mean=mu, sd=sigma)
## [1] 0.7528
```

que confirman los valores del Ejemplo 5.6.2 del libro.

• Ejercicio 23, pág. 41

1. (¡Haz dibujos!) Para el apartado 4 del Ejercicio 21 hacemos:

```
mu = 5
sigma = 3
(k1 = qnorm(0.90, mean=mu, sd=sigma))

## [1] 8.845

(k2 = qnorm(0.95, mean=mu, sd=sigma))

## [1] 9.935
```

Y para el apartado 5:

```
(k1 = qnorm(1 - 0.10, mean=mu, sd=sigma))

## [1] 8.845

(k2 = qnorm(1 - 0.05, mean=mu, sd=sigma))

## [1] 9.935
```

o, lo que es equivalente,

```
(k1 = qnorm(0.10, mean=mu, sd=sigma, lower.tail=FALSE))

## [1] 8.845

(k2 = qnorm(0.05, mean=mu, sd=sigma, lower.tail=FALSE))

## [1] 9.935
```

2. En este caso usamos `qnorm` directamente:

```
mu = -2
sigma = 1/4
(k = qnorm(0.85, mean=mu, sd=sigma))

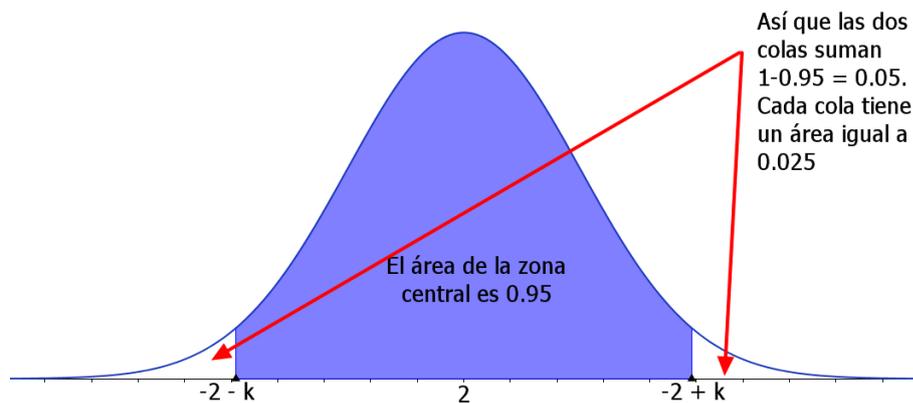
## [1] -1.741
```

3. Ahora tenemos que hacer algún ajuste:

```
(k = qnorm(1 - 0.99, mean=mu, sd=sigma))

## [1] -2.582
```

4. La idea clave de este apartado es que los valores $-2 - k$ y $-2 + k$ son simétricos respecto a la media $\mu = -2$. Por lo tanto el área de la cola izquierda correspondiente al valor $-2 - k$ es igual al área de la cola derecha correspondiente al valor $-2 + k$, como indica la figura.



De esa figura se deduce que el área de la cola **izquierda** correspondiente al valor $-2 + k$ es igual a $0.95 + 0.025 = 0.975$. Es decir, que para localizar $-2 + k$ usamos:

```
(a = qnorm(0.975, mean=mu, sd=sigma))
## [1] -1.51
```

Y entonces (despejando k de $a = -2 + k$):

```
(k = a + 2)
## [1] 0.49
```

Puedes comprobar el resultado usando `pnorm` así:

```
pnorm(-2 + k, mean=mu, sd=sigma) - pnorm(-2 - k, mean=mu, sd=sigma)
## [1] 0.95
```

• **Ejercicio 24, pág. 41**

1. Empezamos generando el vector X con $n = 100000$ elementos.

```
set.seed(2014)
n = 100000
X = rnorm(n, mean=400, sd=15)
```

Ahora, localizamos los elementos de X en el intervalo $(380, 420)$:

```
enIntervalo = (X > 380) & (X < 420)
```

Una vez hecho esto, podemos calcular la fracción del total en ese intervalo:

```
(fraccionEnIntervalo = sum(enIntervalo) / n)
## [1] 0.8169
```

El cálculo teórico de la probabilidad (usando tipificación) sería

```
X1 = 380
X2 = 420
mu = 400
sigma = 15
(Z1 = (X1 - mu) / sigma)
```

```
## [1] -1.333

(Z2 = (X2 - mu) / sigma)

## [1] 1.333

pnorm(Z2) - pnorm(Z1)

## [1] 0.8176
```

Sin tipificar haríamos esto (para obtener el mismo valor):

```
pnorm(X2, mean=mu, sd=sigma) - pnorm(X1, mean=mu, sd=sigma)

## [1] 0.8176
```

Sea como sea, parece que este valor teórico y los resultados de la simulación coinciden razonablemente.

2. Los resultados de la página 178 del libro indican que debería ser:

$$\mu_X = \mu_{X_1} + \mu_{X_2} = 45, \quad \sigma_X = \sqrt{\sigma_{X_1}^2 + \sigma_{X_2}^2} = \sqrt{3^2 + 4^2} = 5.$$

Primero generamos los vectores X1 y X2.

```
set.seed(2014)
n = 100000
X1 = rnorm(n, mean=15, sd=3)
X2 = rnorm(n, mean=30, sd=4)
```

Ahora la suma:

```
X = X1 + X2
```

Y su media y cuasidesviación típica:

```
mean(X)

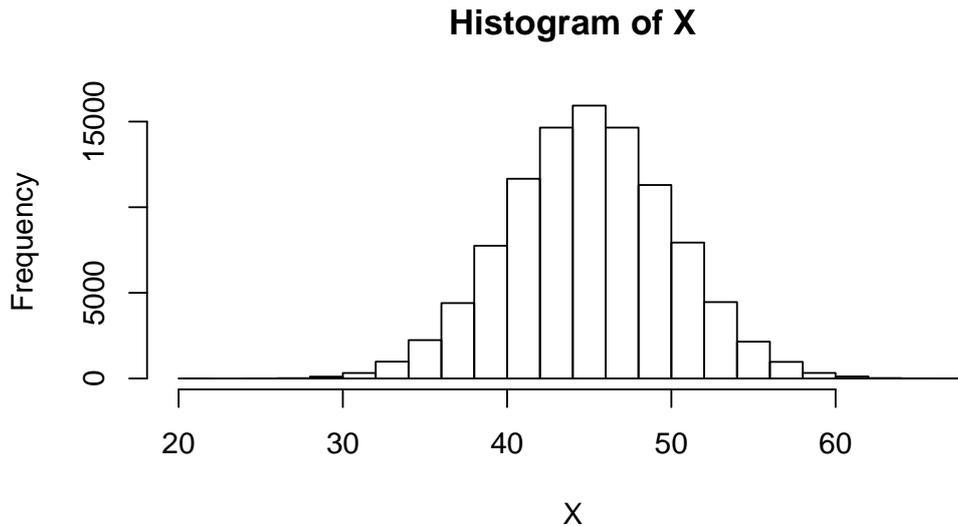
## [1] 44.99

sd(X)

## [1] 5.004
```

Así que la simulación es bastante satisfactoria. Para entender porque hemos usado la cuasidesviación típica en lugar de la desviación típica de X tendremos que avanzar un poco más en el curso. Pero, en cualquier caso, te invitamos a que compruebes que, para un valor de n tan grande, la diferencia entre ambas es casi inapreciable. El histograma aes

```
hist(X)
```



Y, como ves, es lo que esperábamos.

3. Vamos a empezar siguiendo la sugerencia del enunciado para elegir una media y una desviación típica al azar:

```
set.seed(2014)
(mu = signif(runif(1, min = -100, max = 100), 4))

## [1] -42.84

(sigma = signif(runif(1, min = 0.01, max = 100), 4))

## [1] 16.9
```

Una vez hecho esto, generamos una muestra de n valores usando `rnorm`:

```
n = 10000
muestra = rnorm(n, mean = mu, sd=sigma)
```

Y ahora vamos a comprobar cuantos de esos valores pertenecen a los intervalos $\mu \pm k\sigma$, para $k = 1, 2, 3$.

```
(extremosSuperiores = mu + (1:3) * sigma)

## [1] -25.94 -9.04  7.86

(extremosInferiores = mu - (1:3) * sigma)

## [1] -59.74 -76.64 -93.54

unSigma = (muestra < extremosSuperiores[1]) & (muestra > extremosInferiores[1])
dosSigmas = (muestra < extremosSuperiores[2]) & (muestra > extremosInferiores[2])
tresSigmas = (muestra < extremosSuperiores[3]) & (muestra > extremosInferiores[3])

table(unSigma) / n

## unSigma
## FALSE TRUE
## 0.3263 0.6737
```

```

table(dosSigmas) / n

## dosSigmas
## FALSE TRUE
## 0.0481 0.9519

table(tresSigmas) / n

## tresSigmas
## FALSE TRUE
## 0.0031 0.9969

```

Como ves, los porcentajes de valores TRUE se aproximan mucho a lo que predice la regla 68 – 95 – 99.

Puedes probar a ejecutar el código varias veces (recuerda comentar la línea de `set.seed`), para probar con distintas normales y con distintos tamaños muestrales. Prueba por ejemplo con $n = 100$, para ver lo que sucede usando una muestra comparativamente pequeña.

- **Ejercicio 25, pág. 44**

La función, a la que vamos a llamar `f2`, se define mediante:

```
f2 = function(x){ 1 / (pi * (1 + x^2))}
```

- **Ejercicio 26, pág. 44**

Al tratar de evaluar `f` sobre el vector `1:100` se obtiene una advertencia de R, porque cuando la condición de una estructura `if/else` se aplica a un vector, R sólo usa el primer elemento del vector, e ignora todos los restantes elementos.

```
## Warning in if ((x > 0) & (x < 1)) {: la condición tiene longitud > 1 y sólo el primer
elemento será usado

## [1] 0
```

- **Ejercicio 27, pág. 45**

La función es:

```
f = function(x){
  ifelse((x>0) & (x<1), 6 * x * (1 - x), 0)
}
```

Y ahora, como vamos a ver, es posible aplicarla a vectores (como el que aquí fabricamos con `seq`):

```
f(seq(-0.1, 1.1, length.out=11))

## [1] 0.0000 0.1176 0.7224 1.1544 1.4136 1.5000 1.4136 1.1544 0.7224 0.1176
## [11] 0.0000
```

- **Ejercicio 28, pág. 47**

Definimos la función mediante:

```
f = function(x){
  ifelse((1<x)&(x<2), 2* (2 - x), 0)
}
```

Y ahora el resto es como en otros ejemplos que hemos visto:

```
aux_f = function(x){ x * f(x) }
(mu=integrate(aux_f, lower=-Inf, upper=Inf)$value)

## [1] 1.333

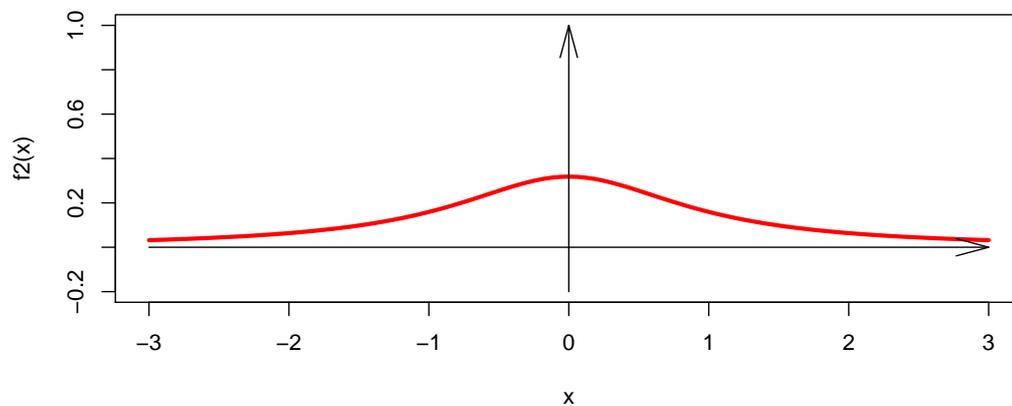
aux2_f = function(x){ (x-mu)^2 * f(x) }
(varianza=integrate(aux2_f, lower=-Inf, upper=Inf)$value )

## [1] 0.05556
```

• Ejercicio 29, pág. 49

1. Vamos a dibujarla en el intervalo $-3 < x < 3$, y representaremos valores de y entre -0.2 y 1 . No te preocupes por el comando `arrows`, que hemos usado para añadir unos ejes de coordenadas “típicos” a la figura.

```
curve(f2, from=-3, to=3, ylim=range(-0.2, 1), col="red", lwd=3)
arrows(c(-3,0), c(0, -0.2), c(3,0), c(0,1), angle=15)
```



2. La definición de la función (a la que llamamos `f3`) usa varios `ifelse` anidados.

```
f3 = function(x){
  ifelse(x<0, 0,
    ifelse(x<=1, 2 * x/3,
      ifelse(x<=2, 0,
        ifelse(x<=3, 4*(3-x)/3, 0))))
}
```

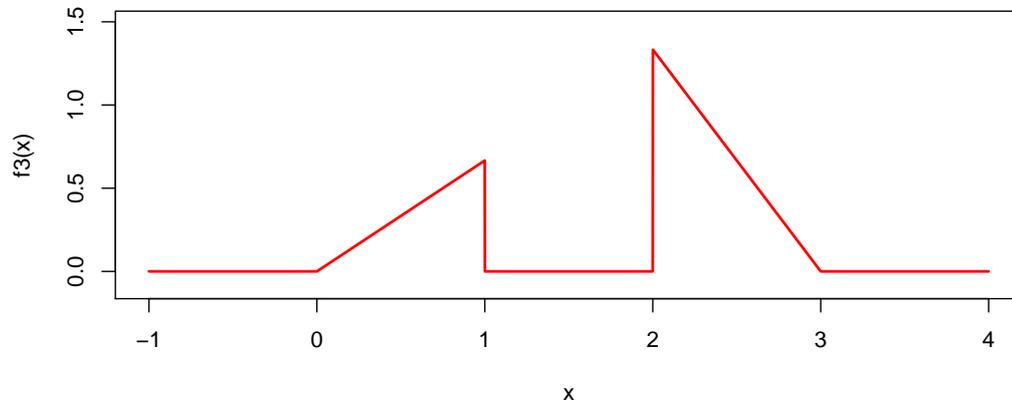
Una vez hecho esto, la gráfica se obtiene con `curve`:

```
xmin = -1
xmax = 4
ymin = -0.1
```

```

ymax = 1.5
curve(f3, from = xmin,to = xmax, ylim=range(ymin, ymax),
      col="red", lwd=2, n = 5000)

```



El argumento `n=5000` sirve para determinar el número de puntos que R usa para dibujar la gráfica.

• **Ejercicio 55, pág. 54**

El código correspondiente se muestra a continuación. Lee los comentarios del código para ver su estructura. La única línea que puede resultarte un poco extraña es la que dice

```
resultados = integer(n)
```

Lo que hacemos en esta línea es crear de antemano el vector en el que se van a guardar los resultados. Después, en cada iteración del bucle `for` el comando

```
resultados[k] = bola
```

se encarga de guardar el resultado del proceso en la posición correspondiente (la número `k`) del vector `resultados`. Esta forma de trabajar resulta un poco más eficiente que si fuéramos concatenando cada resultado al final del vector, con un comando como:

```
resultados = c(resultados, bola)
```

aunque las dos formas producen el mismo resultado.

```

rm(list=ls())
# n es el número de veces que vamos a repetir el proceso.
n = 50000
# Creamos el vector en el que se guardarán los resultados.
resultados = integer(n)

# El proceso ahora depende de si el dado es o no menor que 5
urna = "b"

# El bucle for se encarga de repetir el proceso.
for(k in 1:n){
  # Se usa la urna que corresponde en esta iteración.
  if(urna == "b"){
    #Se usa la urna blanca.
    bola = sample(1:6, 1)

```

```

} else {
  #Se usa la urna negra.
  bola = sample(c(1:5, rep(6,5)), 1)
}
# Se guarda el resultado en el vector.
resultados[k] = bola
# Y se elige, según el resultado, la
# urna que usamos en el siguiente paso.
if(bola == 6){
  urna = "n"
} else {
  urna = "b"
}

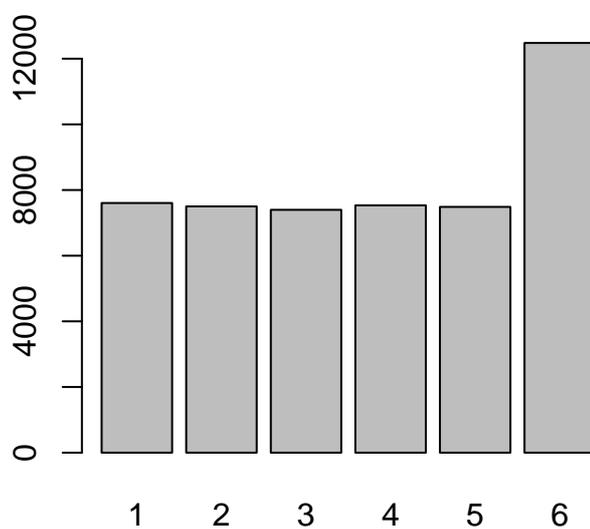
} # Fin del bucle for.

# Al final del bucle, miramos la tabla de frecuencias relativas.
table(resultados) / n

## resultados
##      1      2      3      4      5      6
## 0.1521 0.1500 0.1479 0.1507 0.1497 0.2496

# y la representamos en un diagrama de barras.
barplot(table(resultados))

```



El resultado que se obtiene muestra que el valor 6 es mucho más probable que cualquiera de los restantes cinco valores, que a su vez son todos igual de probables. Este diagrama no se corresponde con ninguna de las distribuciones binomiales que hemos estudiado.

- Ejercicio 56, pág. 54

```

set.seed(2014)
muestra1Cauchy = rcauchy(10000)
mean(muestra1Cauchy)

## [1] -0.7487

muestra2Cauchy = 5 + rcauchy(10000)
mean(muestra2Cauchy)

## [1] 5.906

mean(muestra2Cauchy) - mean(muestra1Cauchy)

## [1] 6.655

```

Ejecuta varias veces el código (sin `set.seed`, claro) para observar lo que sucede.

- **Ejercicio 57, pág. 55**

La integral para $P(A)$ del ejemplo se puede calcular en Wolfram Alpha con:

`integrate (1/pi) * exp(-(x^2+y^2)) dx dy, y=-1 to 1, x=-1 to 1`

Verás que la respuesta involucra a la función `erf`, de la que ya hemos hablado, y que el valor aproximado es 0.7101, como aparece en el libro.

Para calcular $f_X(x)$ del Ejemplo 5.7.3 puedes usar este comando:

`integrate (1/pi) * exp(-(x^2+y^2)) dy, y=-oo to oo`

Fin del Tutorial05. ¡Gracias por la atención!