Cuestiones lectura 3

Cuestión 1:

Para realizar esta cuestión lo primero que habría que realizar es un diagrama de árbol del propio problema. El cual encontraremos en la figura 1.1

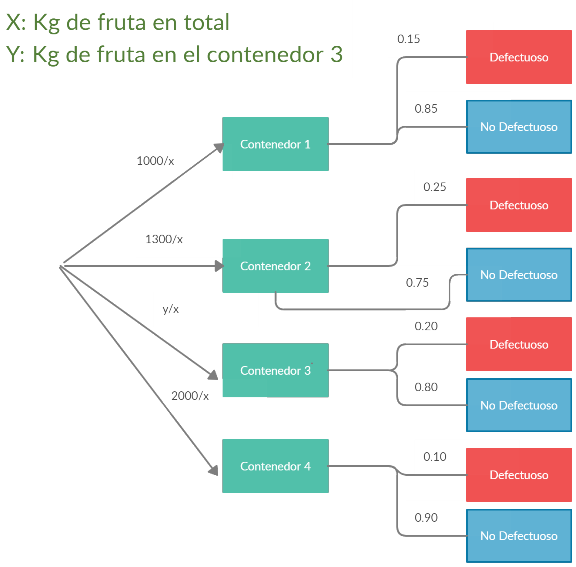


Figure 1: Diagrama de arbol del problema. figura 1.1

Como podemos observar nos faltan dos incógnitas primero, $x = Kg.Totales$ y por otra parte $y = Kg.Contenedor3$. A su vez nos están proporcionando una información clave $P(C_4|D) = 0.22$ es decir la probabilidad de que siendo defectuoso provenga del contenedor 4.

Con esta información ya nos podemos hacer una idea de como será la resolución del problema. Aplicando el teorema de Bayes $P(B_r|A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)}$ podríamos desarrollar esta misma formula substituyendo la intersección por un producto equivalente $P(B_r|A) = \frac{P(A|B_r) * P(B_r)}{\sum_{i=1}^{k} P(A|B_i) * P(B_i)}$ el cual adaptándolo a nuestro problema quedaría $P(C_4|D) = \frac{P(D|C_4) * P(C_4)}{\sum_{i=1}^{4} P(D_i|C_i) * P(C_i)}$.

Al final obtendríamos $0.22 = \frac{0.1 * 200/x}{1000 * 0.15/x + 1300 * 0.25/x + y * 0.2/x + 200/x}$ y resolviendo podemos eliminar las x quedándonos tan solo nuestra incógnita y.$0.22 = \frac{0.1 * 200}{1000 * 0.15 + 1300 * 0.25 + y * 0.2 + 200}$

Resolviendo en R podemos observar que los kg del contenedor 3 son:

```
p=1000*0.15+1300*0.25+200;
x=12875/11; x
```

```
## [1] 1170.455
```

Cuestión 2:

En esta cuestión tendremos que operar con conjuntos de probabilidad. empecemos realizando $P(E_1 \cap E_2 \cap E_3)$ para ello bastará con multiplicar todas las diferentes probabilidades entre si.

```
0.8*0.65*0.75
```

```
## [1] 0.39
```

$P(E_1 \cup E_2)$ En este caso tendremos que sumar ambas probabilidades y restarle la intersección debido que esta se repite.

```
0.8+0.65-0.8*0.65
```

```
## [1] 0.93
```

$P(E_1 \cup E_2 \cup E_3)$ Este es el mismo caso anterior pero de forma generalizada.
Sumamos las variables y restamos todas las intersecciones entre las diferentes probabilidades y por último

```
0.8+0.65+0.75-0.8*0.65-0.8*0.75-0.75*0.65+0.39
```

```
## [1] 0.9825
```

$P(E_1 \cup \neg E_2 \cup \neg E_3)$ Sería el apartado anterior pero usando las probabilidades inversas

```
0.8+0.35+0.25-0.8*0.35-0.8*0.25-0.35*0.25+0.8*0.35*0.25
```

```
## [1] 0.9025
```

Cuestión 3:

La cuestión 3 es muy parecida a la 1. Nos proporcionan unos datos con los que poder confeccionar un problema usando el T. de Bayes. Primero nos proporcionan la cantidad de población que tiene la enfermedad $P(E) = 0.025$, en segundo lugar $P(T|E) = 0.95$ y por último $P(T|!E) = 0.1$.

Nos solicitan la probabilidad de que si un paciente llega al servicio y da positivo este enfermo, es decir, $P(E|T)$

Usando sencillamente la formula siguiente, que es un aplicación del T. de Bayes $P(E|T) = P(T|E) * P(E)/P(T)$ obtendríamos el resultado. Siendo este:

```
0.95*0.025/(0.95*0.985+0.1*0.025)
```

```
## [1] 0.02531308
```

Cuestión 4:

Tenemos dos conjuntos aquellos formados por pares de dos letras iguales ej ABBA y por otra parte aquellos formados por 3 letras iguales y 1 diferente. Podemos observar que ambos problemas pueden ser analizados como variaciones pues importa el orden, no se cogen todos los elementos, y estos no se pueden repetir. Por lo tanto seria tan sencillo como en el a realizar variacione 4 en 1 multiplicado por n (27) y n-1. Y el suceso b seria coger 2 de cuatro y mutiplicar por n y por n-1. Esto es debido a que en uno tenemos pares de dos en dos es decir cogemos de 4 dos y en otro solo tenemos una variacion de un espacio.

```
#a

choose(4, 1) * 27 *26
```

```
## [1] 2808
```

```r
#b

choose(4, 2) * 27 * 26
```

```
## [1] 4212
```

Cuestión 5:

En el enunciado nos comentan que existen de un conjunto de 100 elementos 5 defectuosos y de esos 100 cogemos 50. El planteamiento del problema lo he realizado de la siguiente manera la probabilidad de que tengamos un defectuoso es de 5/100 y cogemos los elementos de 50 en 50 por lo tanto podemos usar una binomial para calcular el resultado.

```r
pbinom(1, 50, 0.05)
```

```
## [1] 0.2794318
```

Cuestión 6:

Para abordar la cuestión tendremos que hablar de la probabilidad total es decir la probabilidad de lesión P(L) sera 1 pues hablamos de un conjunto donde si o si alguien ha resultado lesionado. Y P(L) seria la suma de P(T u H) + P(!(P U H)).

Probabilidad de que uno acuda al hospital es $P(H) = 0.10$, la probabilidad de que acuda al trabajo el siguiente día $P(T) = 0.15$ y ambas $P(T \cap H) = 0.02$

Cuestiones Lectura 4

Cuestión 1:

Lo primero que realizaremos sera establecer como vamos a trabajar. Primero crearemos dos listas con las ganancias y las probabilidades. Luego las agregaremos a un data.frame para poder trabajar mas fácilmente.

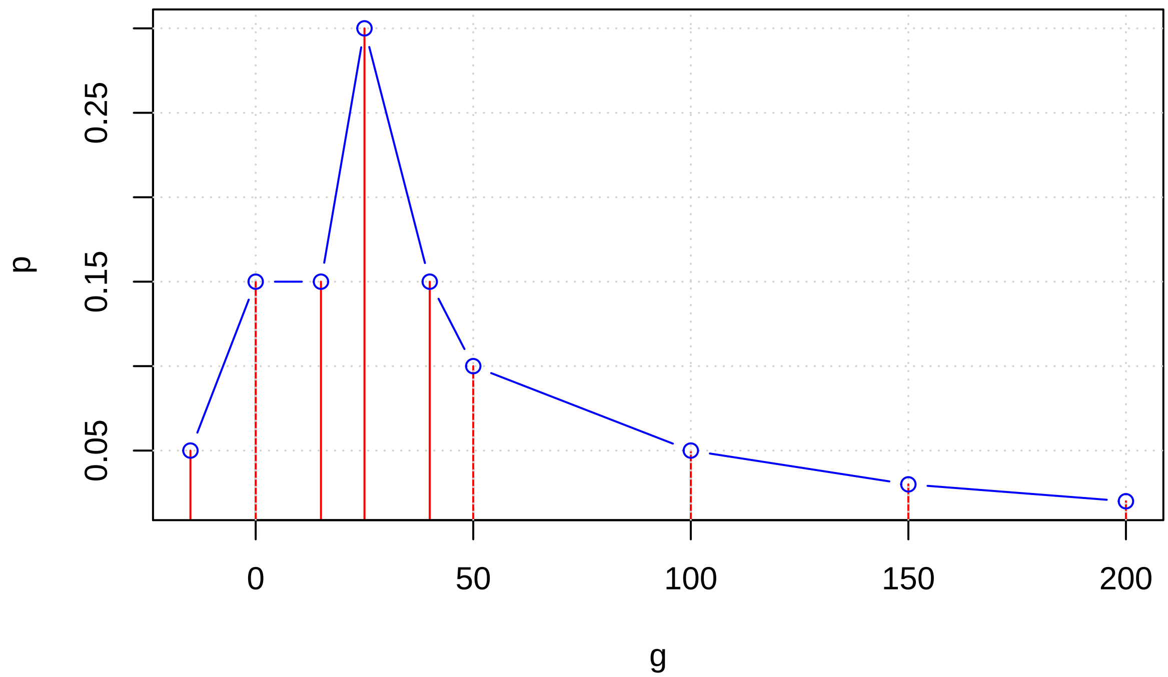
```
g = c(-15, 0, 15, 25, 40, 50, 100, 150, 200)
p = c(0.05, 0.15, 0.15, 0.30, 0.15, 0.10, 0.05, 0.03, 0.02)

p_frame = data.frame("ganancias" = g, "probabilidad" = p); p_frame
```

```
##   ganancias probabilidad
## 1       -15         0.05
## 2         0         0.15
## 3        15         0.15
## 4        25         0.30
## 5        40         0.15
## 6        50         0.10
## 7       100         0.05
## 8       150         0.03
## 9       200         0.02
```

A continuación previsualizaremos la información para tener una idea clara de con que estamos trabajando.

```
plot(g, p, type='h', col= "red")
grid()
points(g, p, type="b", col="blue")
```



Ya por último calcularemos el valor esperado sumando todas las ganancias por la probabilidad, y con este valor calcularemos la varianza como la suma de las ganancias - el valor esperado elevado a la probabilidad por 2.

```
f_acum = cumsum(p)

p_frame$p_acum = f_acum; p_frame
```

```
##   ganancias probabilidad p_acum
## 1       -15         0.05   0.05
## 2         0         0.15   0.20
## 3        15         0.15   0.35
## 4        25         0.30   0.65
## 5        40         0.15   0.80
## 6        50         0.10   0.90
## 7       100         0.05   0.95
## 8       150         0.03   0.98
## 9       200         0.02   1.00
```

```
valor_esperado = sum(p_frame$ganancias * p_frame$probabilidad); valor_esperado
```

```
## [1] 33.5
```

```
varianza = sum((p_frame$ganancias-valor_esperado)^ 2* p_frame$probabilidad); varianza
```

```
## [1] 1575.25
```

Cuestión 2

#a)

Para calcular la distribución de probabilidad haremos uso de la función choose y de un bucle for que nos permitirá ir calculando la probabilidad de cada opcián.

Quedándonos con una tabla que nos muestra esta distribución.

```
library(MASS)
m = matrix(0:15, ncol=4, nrow=4)
print("Maneras de coger las frutas ")
```

```
## [1] "Maneras de coger las frutas "
```

```
print(choose(8,4))
```

```
## [1] 70
```

```
for(x in 1:4){
  for(y in 1:4){
    m[y, x] = choose(3, x - 1) * choose(2, y - 1) * choose(3, -(x-1+y-1)+4)
  }
}

m = fractions(m/choose(8,4))
m
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    0 3/70 9/70 3/70
## [2,] 1/35 9/35 9/35 1/35
## [3,] 3/70 9/70 3/70    0
## [4,]    0    0    0    0
```

b)

Este apartado consiste en sumar todas aquellas probabilidades donde la suma de las naranjas y manzanas cogidas no supere el 2

```
m[1,1] + m[1,2] + m[2,1] + m[2,2]+ m[1,3] + m[3,1]
```

```
## [1] 1/2
```

c)

Por ultimo calcularemos la media de x y le restaremos la media de y.

```
var = c(0,1,2, 3)
valor_esperado_x = 0

for(x in 1:4){
  for(y in 1:4){
    valor_esperado_x = valor_esperado_x + (var[x] * (m[x,y]));
  }
}

valor_esperado_x
```

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

```r
valor_esperado_y = 0

for(x in 1:4){
  for(y in 1:4){
    valor_esperado_y = valor_esperado_y + var[y] * (m[x,y]);
  }
}

valor_esperado_y
```

```
## [1] 3/2
```

```r
valor_esperado_x - valor_esperado_y
```

```
## [1] -1/2
```

Cuestión 3

##a y b

Para calcular la distribución marginal en X basta con sumar las filas y lo mismo para la Y pero con las columnas dando como resultado dos vectores cada uno con 3 valores.

```r
p_frame = data.frame("1" = c(0.05, 0.05, 0.00), "2"=c(0.05, 0.10, 0.20), "3"=c(0.10, 0.35, 0.10))
rownames(p_frame)=c('Y1','Y3', 'Y5' ); p_frame
```

```
##      X1   X2   X3
## Y1 0.05 0.05 0.10
## Y3 0.05 0.10 0.35
## Y5 0.00 0.20 0.10
```

```r
m = matrix(c(0.05, 0.05, 0, 0.05, 0.10, 0.20, 0.10, 0.35, 0.10), nrow=3, ncol=3)


distribucion_m_x = c(sum(p_frame[1,]), sum(p_frame[2,]), sum(p_frame[3,])); distribucion_m_x
```

```
## [1] 0.2 0.5 0.3
```

```r
distribucion_m_y = c(sum(p_frame$X1), sum(p_frame$X2), sum(p_frame$X3)); distribucion_m_y
```

```
## [1] 0.10 0.35 0.55
```

c

La probabilidad en Y = 3 y X = 2 se puede ver de manera sencilla realizando el siguiente cálculo P(Y | X) = P(X intersección Y)/P(X)

```r
p_3_2 = 0.10/0.35; p_3_2
```

```
## [1] 0.2857143
```

d

Para calcular las medias y esperanza bastará con hacer uso del sumatorio de x * f(x,y), y * f(x,y), x * y * f(x,y)

```r
var = c(1,3,5)

valor_esperado_x = 0

for(x in 1:3){
  for(y in 1:3){
    valor_esperado_x = valor_esperado_x + x * (p_frame[x,y]);
  }
}

valor_esperado_x
```

```
## [1] 2.1
```

```r
var = c(1,3,5)

valor_esperado_y = 0

for(x in 1:3){
  for(y in 1:3){
    valor_esperado_y = valor_esperado_y + var[y] * (p_frame[x,y]);
  }
}

valor_esperado_y
```

```
## [1] 3.9
```

```r
valor_e_xy = 0

for(x in 1:3){
  for(y in 1:3){
    valor_e_xy = valor_e_xy + var[y] * x * (p_frame[x,y]);
  }
}

valor_e_xy
```

```
## [1] 8.2
```

e (Preguntar)

El valor esperado se calcula realizando un sumatorio en x del sumatorio en y de g(x, y) en este caso XY^2 por f(x, y)

```r
var = c(1,3,5)

valor_esperado = 0

for(x in 1:3){
  for(y in 1:3){
    valor_esperado = valor_esperado + (x*var[y]^2) * (p_frame[x,y]);
  }
}

valor_esperado
```

```
## [1] 35.3
```

Cuestión 4

a)

Como podemos observar estamos ante una distribucion binomial puesto que o bien pueden estar desechadas las latas o bien pueden no estarlo. Por otra parte para calcular la probabilidad de que la lata este incompleta o no basta con dividir 150/ 5000 y para calcular que al menos exista una lata incompleta, es decir, que haya una o mas incompletas basta con restar 1 a la probabilidad de que no haya ninguna incompleta.

```r
p_incom = 150/5000; p_incom
```

```
## [1] 0.03
```

```r
#probabilidad de que haya almenos una
1 - dbinom(0, 15, p_incom)
```

```
## [1] 0.3667488
```

b)

La media de latas a desechar se puede calcular tan sencillamente como multiplicar la cantidad de latas que inspeccionaremos por el porcentaje que hay incompletas. A su vez otra forma de calcularlo es sumando las probabilidades de cada lata incompleta * el numero de latas incompletas, es decir sumatorio x*f(x)

```r
#media de latas para desechar
latas_esperadas = 15*p_incom; latas_esperadas
```

```
## [1] 0.45
```

```r
sum(dbinom(0:15, 15, p_incom) * 0:15)
```

```
## [1] 0.45
```

c)

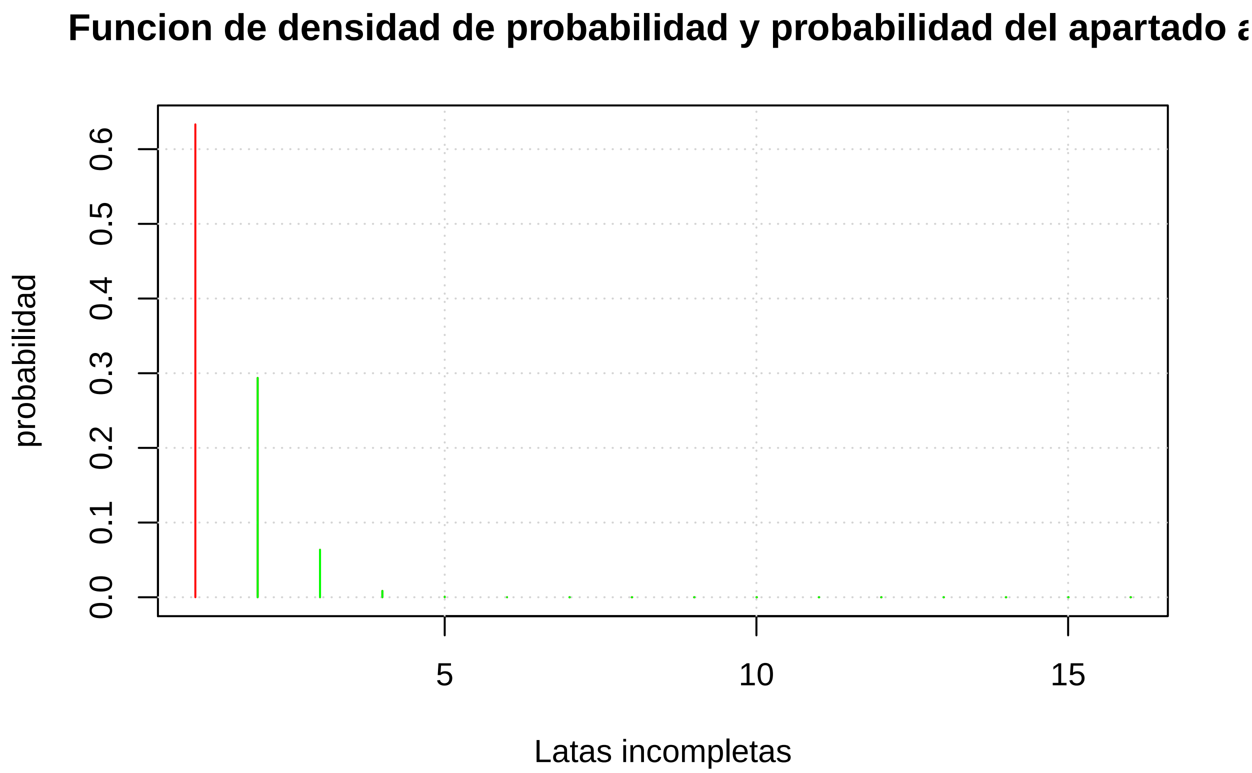
En este caso calcularemos una puntuación puntual usando dbinom.

```r
dbinom(3, 15, p_incom)
```

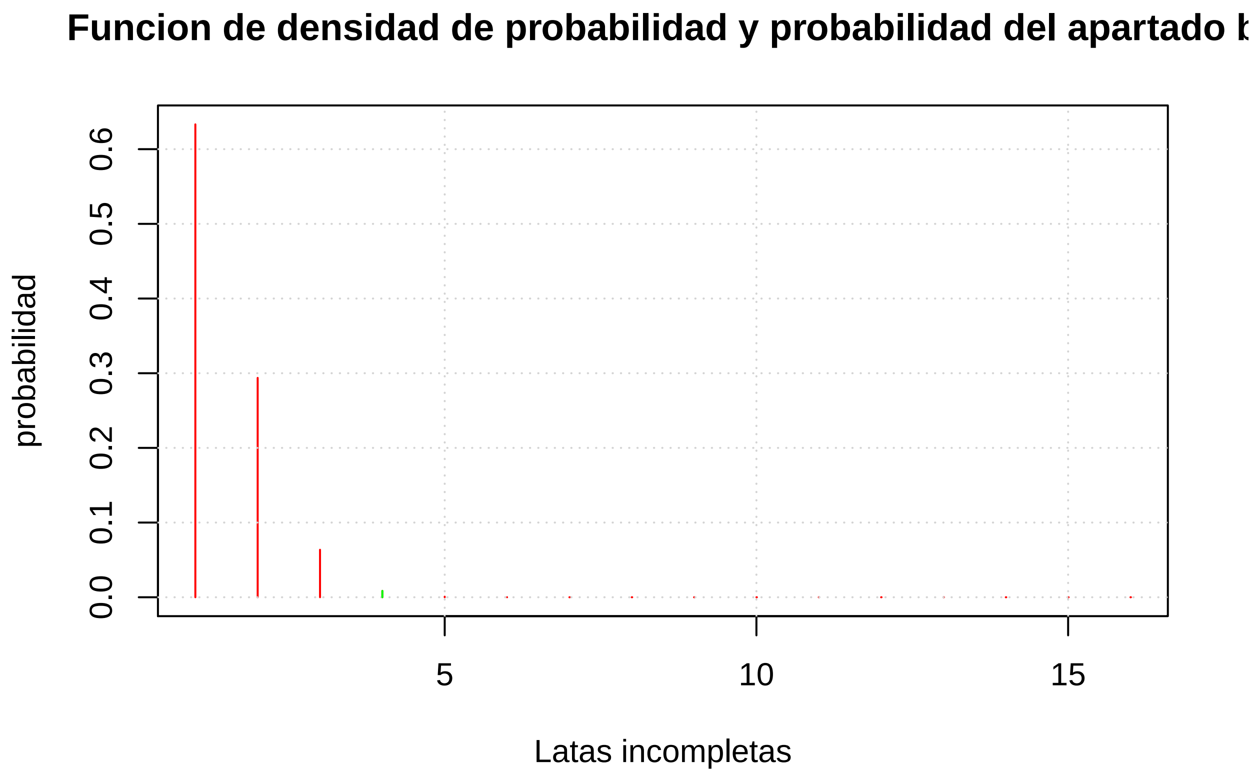
```
## [1] 0.008523853
```

d)

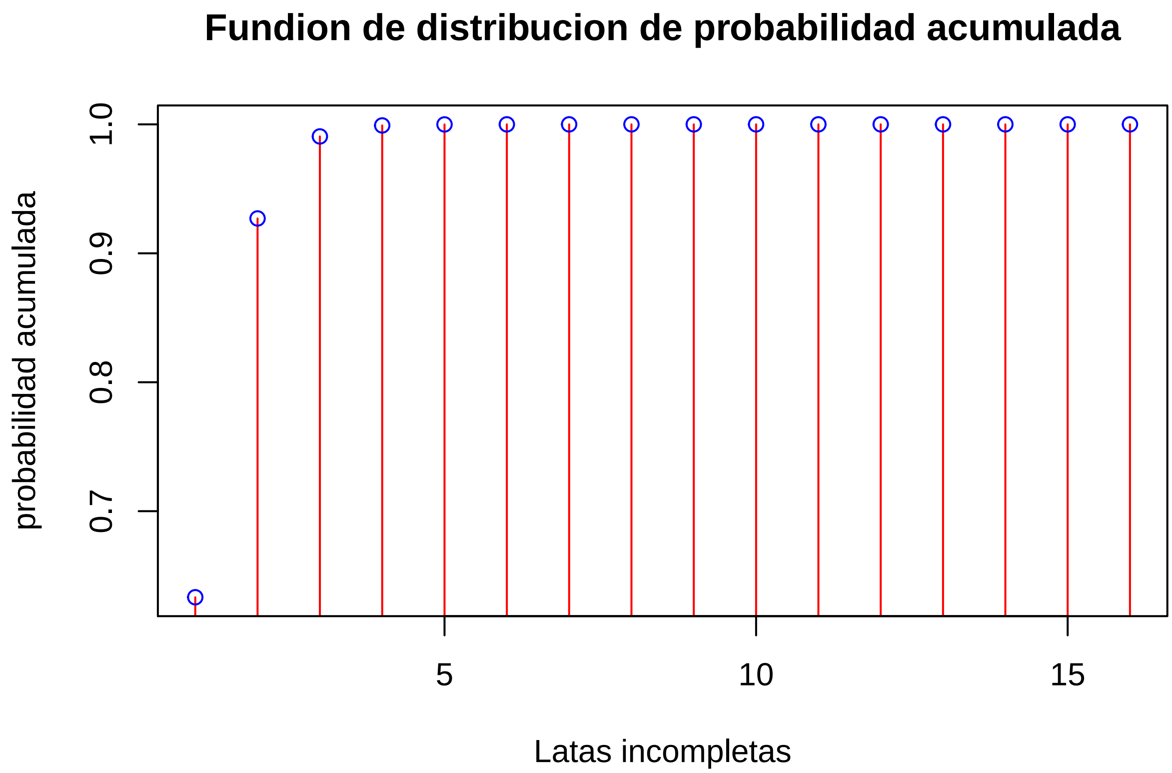
```r
plot(seq(0:15), dbinom(0:15, 15, p_incom), type='h', col="red", main ="Funcion de densidad de probabili
grid()
points(2:16, dbinom(1:15, 15, p_incom), col='green', type='h')
```




```
plot(seq(0:15), dbinom(0:15, 15, p_incom), type='h', col="red", main ="Funcion de densidad de probabili
grid()
points(4, dbinom(3, 15, p_incom), col='green', type='h')
```




```
plot(seq(0:15), pbinom(0:15, 15, p_incom), type='h', col="red", main ="Fundion de distribucion de probal
points(seq(0:15), pbinom(0:15, 15, p_incom),  col="blue")
```



Cuestión 5

Esta cuestión puede ser planteada con múltiples tipos de distribuciones bien una binomial considerando n = 360 y p = 0.08, una aproximación normal o bien una exponencial negativa.

En mi caso he decidido resolverlo usando una distribución de probabilidad binomial.

a)

La probabilidad de este apartado es binomial puesto que de 365 días puede haber o no accidentes o no (éxito o fracaso). En el apartado a nos piden la probabilidad que de 365 días haya un accidente esta se puede calcular fácilmente usando dbinom.

```
dbinom(1, 365, 0.08)
```

```
## [1] 1.923765e-12
```

b)

Nos piden la probabilidad de que como máximo ocurran 3 accidentes por lo tanto tendremos que calcular la probabilidad acumulada hasta 3.

```
pbinom(3, 365, 0.08)
```

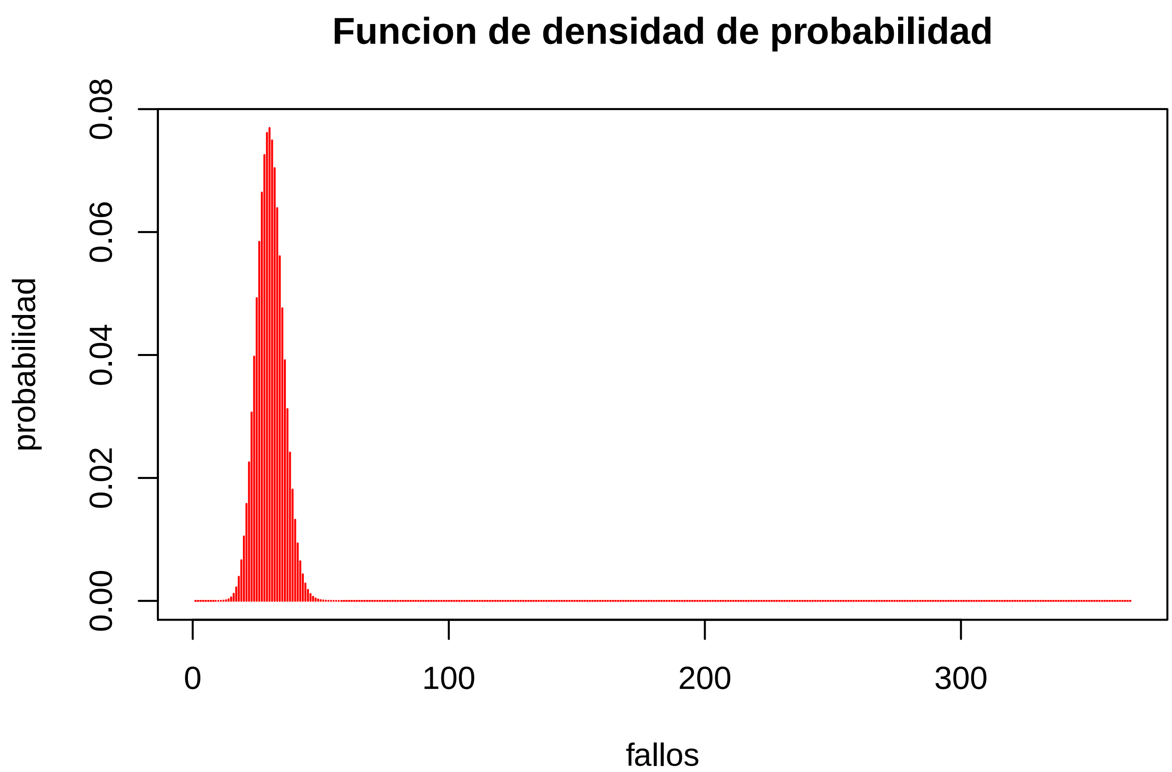
```
## [1] 3.527715e-10
```

c)

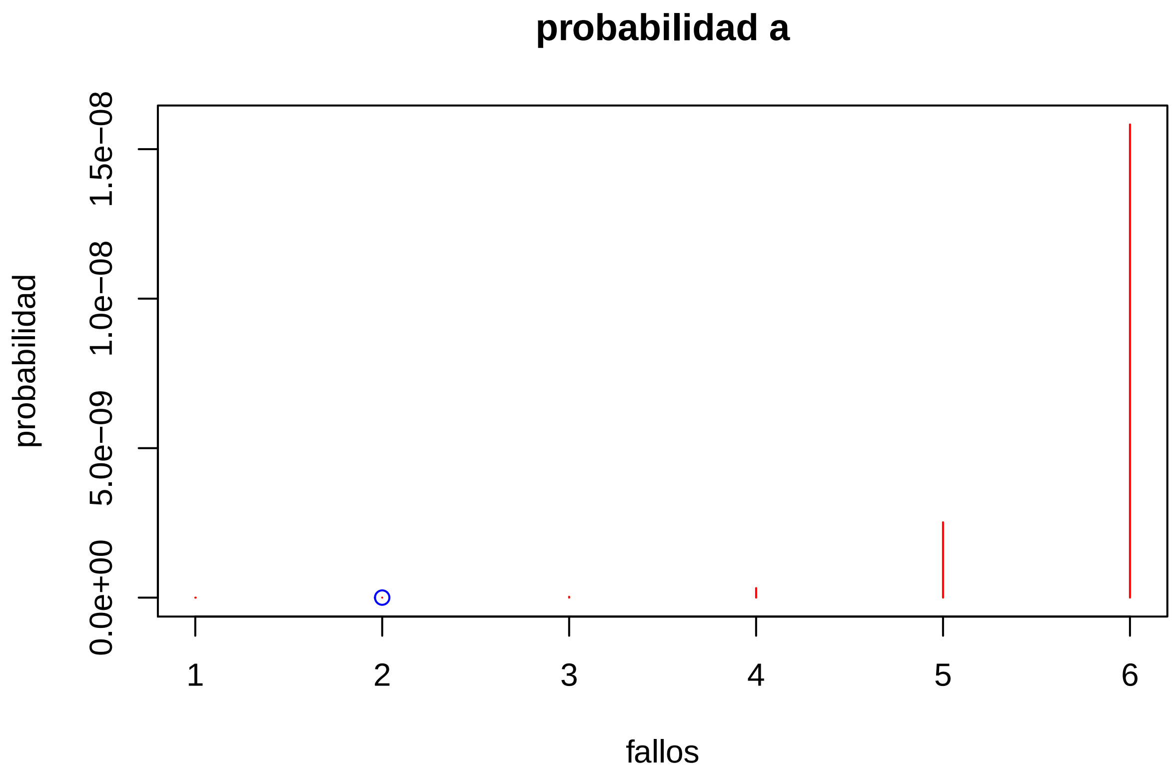
Como he indicado anteriormente la probabilidad subyacente es la binomial, puesto que existen casos de éxito y fracaso y es una probabilidad discreta. Los parámetros serían la probabilidad de éxito, el tamaño de la muestra, y los éxitos.

d)

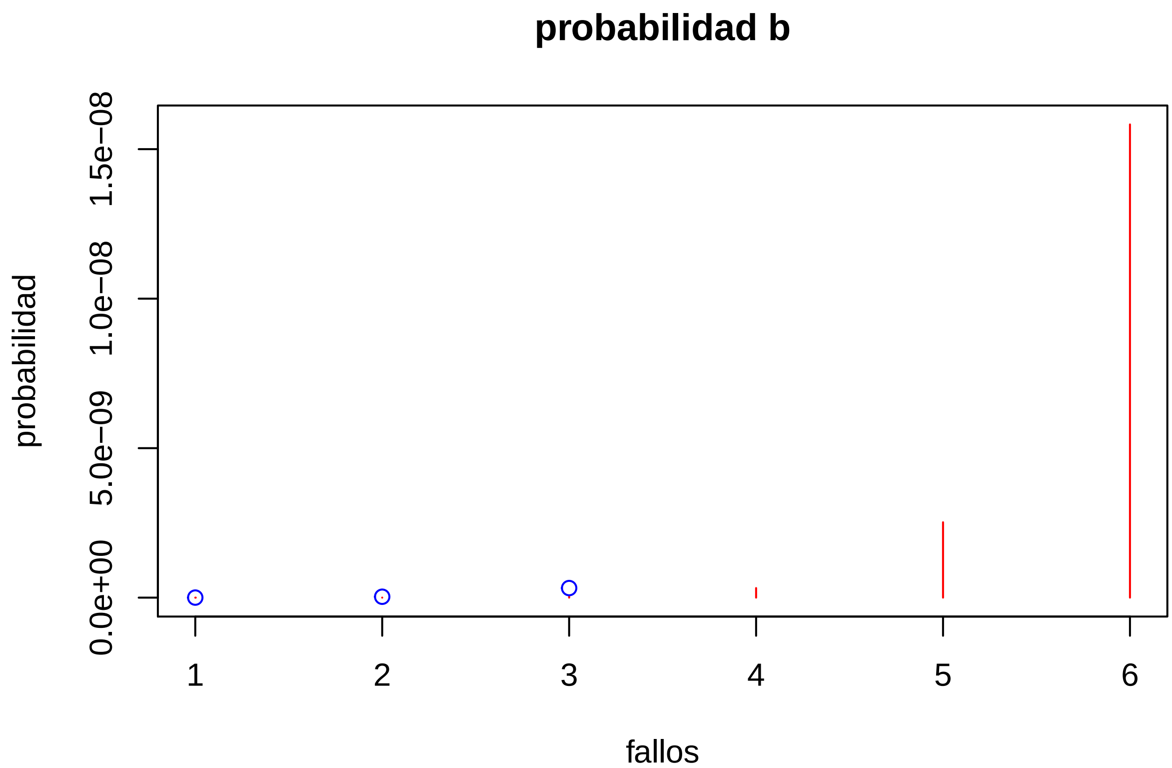
```
plot(seq(0:365), dbinom(0:365, 365, 0.08), type='h', col="red", main ="Funcion de densidad de probabili
```




```
plot(seq(0:5), dbinom(0:5, 365, 0.08), type='h', col="red", main ="probabilidad a", ylab="probabilidad"
points(2, dbinom(1, 365, 0.08),  col="blue")
```

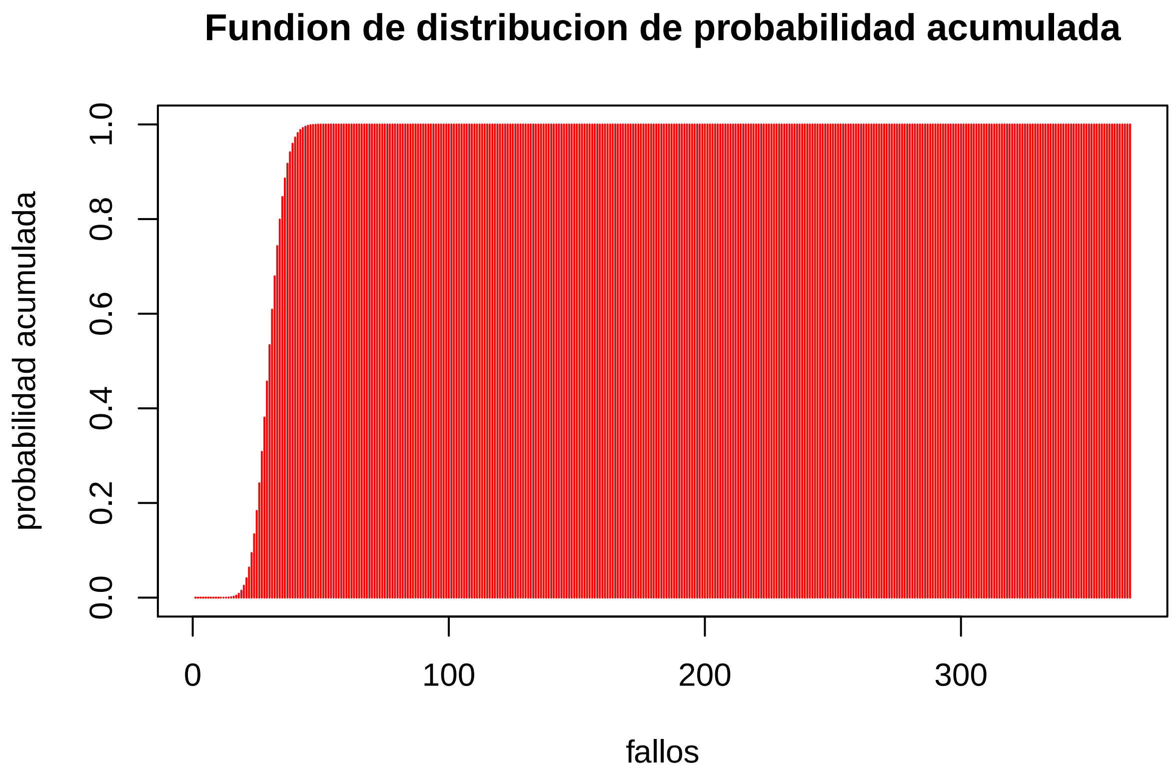



```r
plot(seq(0:5), dbinom(0:5, 365, 0.08), type='h', col="red", main ="probabilidad b", ylab="probabilidad"
points(0:3, dbinom(0:3, 365, 0.08),  col="blue")
```




```
plot(seq(0:365), pbinom(0:365, 365, 0.08), type='h', col="red", main ="Fundion de distribucion de probal
points(seq(0:365), pbinom(0:365, 15, 365),  col="blue")
```

```
## Warning in pbinom(0:365, 15, 365): NaNs produced
```



Cuestión 6

a)

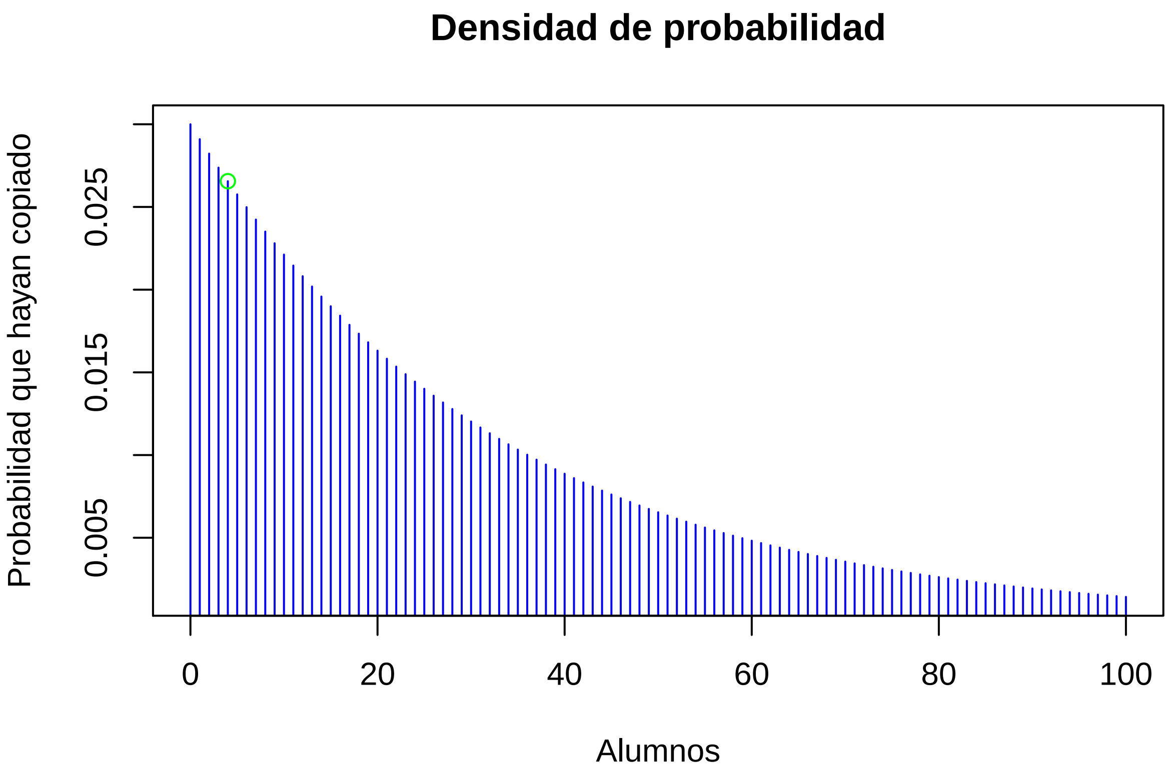
```
p = 3/100;p
```

```
## [1] 0.03
```

```
dgeom(5-1, p)
```

```
## [1] 0.02655878
```

```
plot(0:100, dgeom(0:100, p), type="h", col="blue", main="Densidad de probabilidad", xlab="Alumnos", ylab=
       "Probabilidad que hayan copiado")
points(5-1, dgeom(5-1, p), col="green", main="Densidad de probabilidad", xlab="Alumnos", ylab=
       "Probabilidad que hayan copiado")
```



b)

La distribución que estamos tratando es una hipergeométrica, una distribución discreta relacionada con muestreos aleatorios sin reemplazo. Los parámetros son N la población de elementos, d pertenecientes a la categoría A y N-d los pertenecientes a la B. Y esta calcula la probabilidad de obtener x elementos de la categoría A en una muestra sin reemplazo de n elementos de la población original.

Cuestiones Lectura 5

Cuestión 1

a)

Para ello tenemos que calcular la probabilidad normal acumulada hasta 125 y restarle la acumulada hasta 100. Finalmente esta probabilidad la multiplicaremos por el número de alumnos para obtener el dato que nos piden.

```
550 * (pnorm(125,mean= 118, sd= 12) - pnorm( 100,mean= 118,sd= 12))
```

```
## [1] 359.3471
```

b)

En este caso nos piden por encima de 120 por ello debemos calcular la inversa de la acumulada hasta 120 y multiplicarla por 550 alumnos.

```
550 * (1 - pnorm(120, 118, 12))
```

```
## [1] 238.5989
```

c)

Lo mismo con este. Calculamos la acumulada hasta 95 y la multiplicamos por los 500 alumnos con los que contamos.

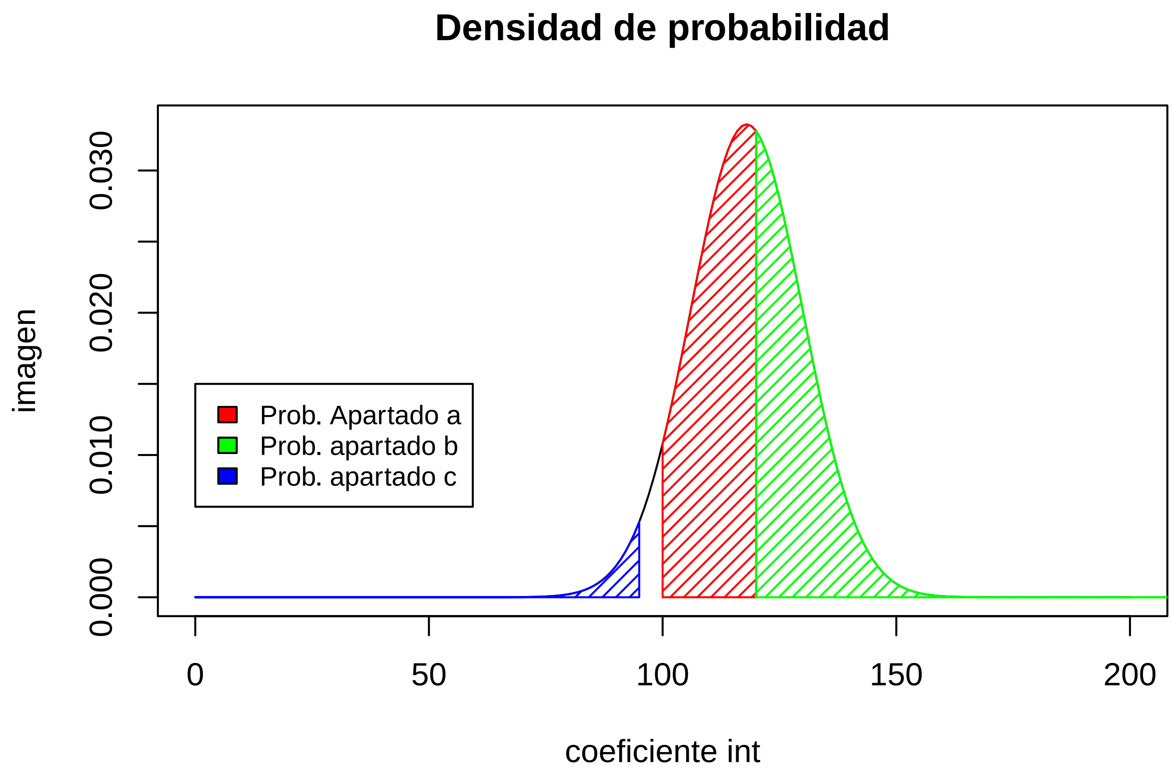
```
550 * pnorm(95, 118, 12)
```

```
## [1] 15.20208
```

d)

Ahora mostraremos un gráfico con la densidad de probabilidad y las diferentes probabilidades calculadas

```
z <- c(100, seq(100,120), 120)
p <- c(0, dnorm(seq(100,120), 118, 12),0)
plot(-0:200, dnorm(0:200, 118, 12), type="l", main="Densidad de probabilidad", xlab='coeficiente int', y
polygon(z, p, density=c(20, 20), col="red")
z <- c(120, seq(120,5000), 5000)
p <- c(0, dnorm(seq(120,5000), 118, 12),0)
polygon(z, p, density=c(20, 20), col="green")
z <- c(0, seq(0,95), 95)
p <- c(0, dnorm(seq(0,95), 118, 12),0)
polygon(z, p, density=c(20, 20), col="blue")
legend(0, 0.015, legend=c("Prob. Apartado a", "Prob. apartado b", "Prob. apartado c"),
       col=c("red", "blue"), fill=c('red', 'green','blue'), cex=0.8)
```



Cuestión 2

a)

```
qexp(0.99 , 1)
```

```
## [1] 4.60517
```

b)

```
pexp(0.01, 4)
```

```
## [1] 0.03921056
```

c)

```
1 - pexp(0.01, 200)
```

```
## [1] 0.1353353
```

Cuestión 3

a)

Para ello realizaremos la integral de x*f(x) de -40 a 60.

```
funcion = function(x) { x * (6e-6*(-x^2+20*x+2400))}
media <- integrate(funcion, -40, 60)
media
```

```
## 10 with absolute error < 2.3e-13
```

b)

Este apartado haremos uso de la integral de (x+media) * f(x) para calcular la sd

```
#funcion_2 = function(x) {(x+18.8)^2*((6e-6*(-x^2+20*x+2400)))}
funcion_2 = function(x) {(x+10)^2*((6e-6*(-x^2+20*x+2400)))}
otro<- integrate(funcion_2, -40, 60)
otro
```

```
## 900 with absolute error < 1e-11
```

c)

Para ello integraremos f(x) entre 10 y 20, ya que recordemos que en este tipo de d. de probabilidad no existen las probabilidades puntuales.

```
funcion = function(x) {6e-6*(-x^2+20*x+2400)}
valor = integrate(funcion, 10, 20)
valor
```

```
## 0.148 with absolute error < 1.6e-15
```

d)

Repetiremos el mismo proceso entre -40 y 0 + 5 minutos de cortesía

```
valor = integrate(funcion, -40, 5)
valor
```

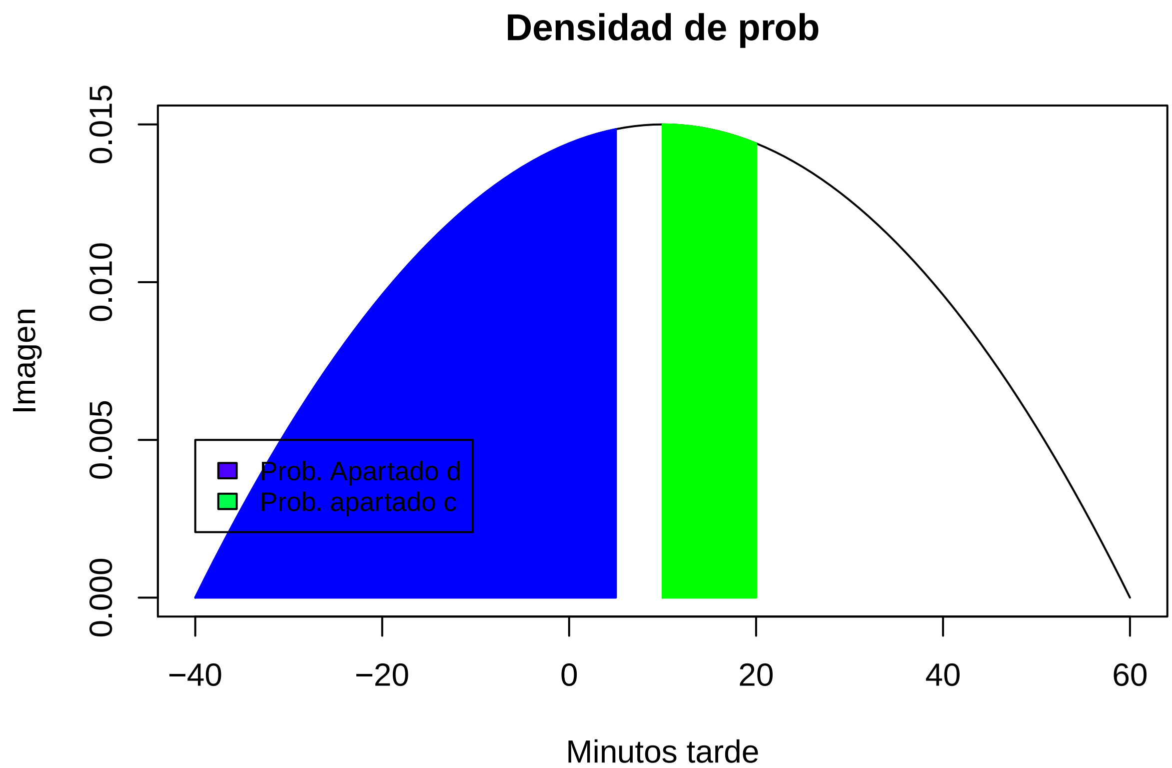
```
## 0.42525 with absolute error < 4.7e-15
```

e)

Ya para terminar mostraremos los resultados obtenidos haciendo uso de las funcionalidades de R

```
z <- c(10, seq(10,20), 20)
p <- c(0, funcion(seq(10,20)),0)

plot(-40:60, funcion(-40:60), type="l", main="Densidad de prob", ylab='Imagen', xlab='Minutos tarde')
polygon(z, p, density=c(-40, 5), col="green")
z <- c(-40, seq(-40,5), 5)
p <- c(0, funcion(seq(-40,5)),0)
polygon(z, p, density=c(-40, 5), col="blue")
legend(-40, 0.005, legend=c("Prob. Apartado d", "Prob. apartado c"),
       col=c("red", "blue"), fill=topo.colors(3), cex=0.8)
```



Cuestión 4

a)

Este apartado se puede realizar sencillamente calculando la probabilidad acumulada hasta 230 y restandole uno para obtener la inversa

```
1- pnorm(230,  sd= 30, mean = 170)
```

```
## [1] 0.02275013
```

b)

Para este apartado usaremos la probabilidad calculada anteriormente y realizaremos una binomial con esta contando como éxito el colesterol por encima de 230.

```
p <- 1- pnorm(230,  sd= 30, mean = 170)
1 - pbinom(8, 180, p)
```

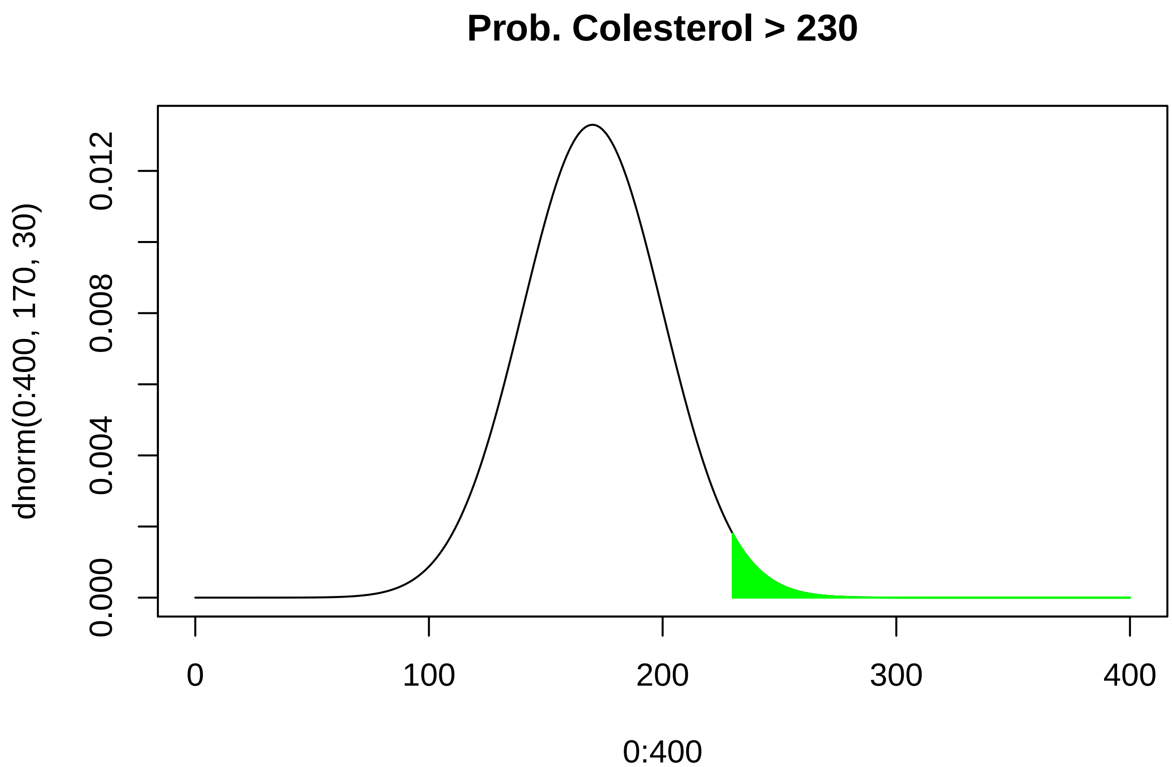
```
## [1] 0.02287882
```

c)

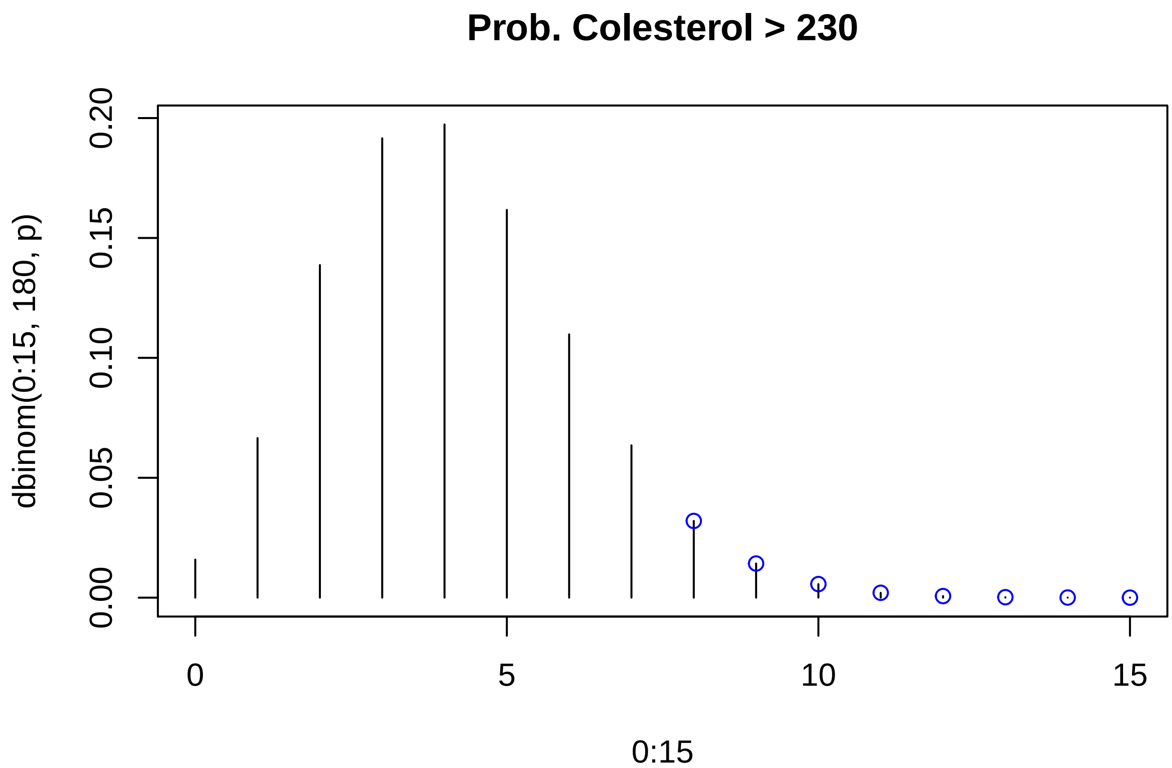
Para terminar mostraremos las distribuciones calculadas anteriormente con sus respectivas probabilidades

```
z <- c(230, seq(230,400), 400)
p <- c(0, dnorm(seq(230,400), 170, 30),0)
```

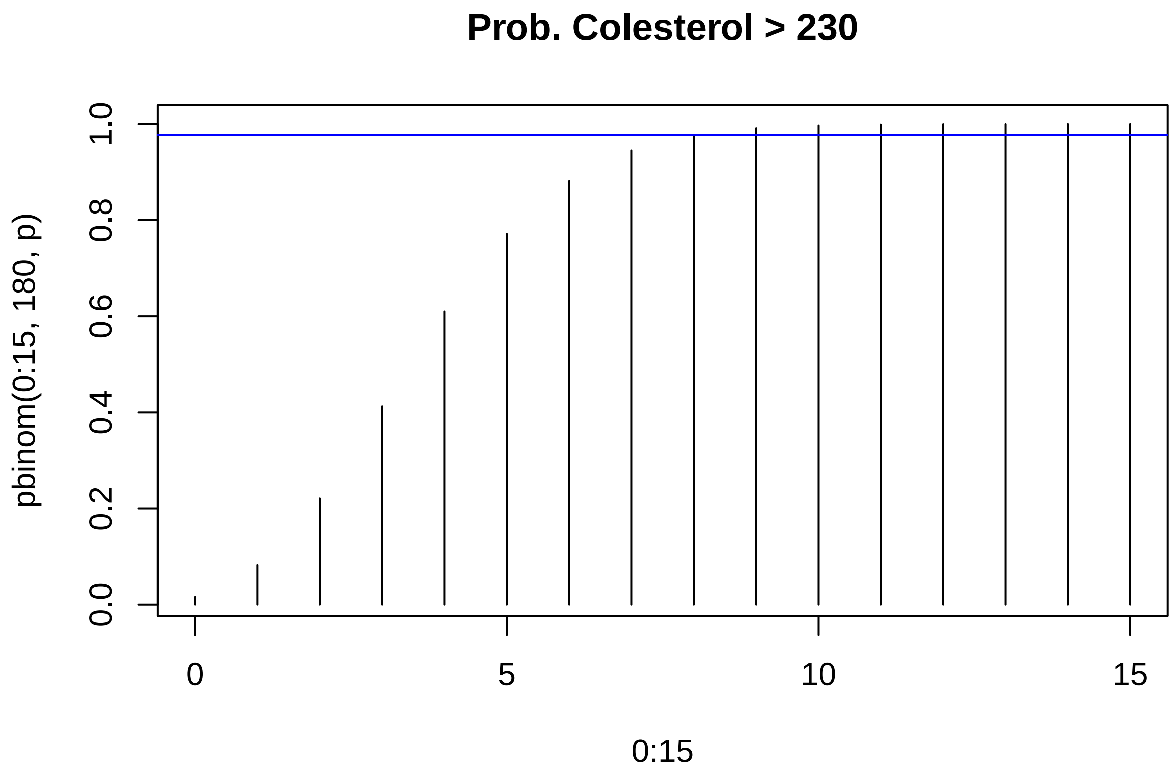
```
plot(0:400, dnorm(0:400, 170, 30), type="l", main="Prob. Colesterol > 230")
polygon(z, p, density=c(-40, 5), col="green")
```




```
p <- 1- pnorm(230,  sd= 30, mean = 170)
plot(0:15, dbinom(0:15,180, p), type="h", main="Prob. Colesterol > 230")
points(8:400, dbinom(8:400, 180, p),col="blue")
```




```r
p <- 1- pnorm(230,  sd= 30, mean = 170)
plot(0:15, pbinom(0:15,180, p), type="h", main="Prob. Colesterol > 230")
abline(pbinom(8, 180, p),0,col="blue")
```



Cuestión 5

a)

Para calcular alpha y beta bastará con dividir la varianza entre la media para calcular beta y a lo calcularemos dividindo la media entre beta calculada anteriormente

```
a = 6/2;a
```

```
## [1] 3
```

```
b = 12/6; b
```

```
## [1] 2
```

b)

Una vez con los datos para empezar a calcular las probabilidades será tan sencillo como usar la función pgamma de R hasta 12 y restarle 1 para calcular la inversa

```
1 - pgamma(12 , shape=a, scale=b)
```

```
## [1] 0.0619688
```

c)

Para calcular esta probabilidad calcularemos la acumulada hasta 4 con los datos anteriores

```
pgamma(4 , shape=a, scale=b)
```

```
## [1] 0.3233236
```

d)

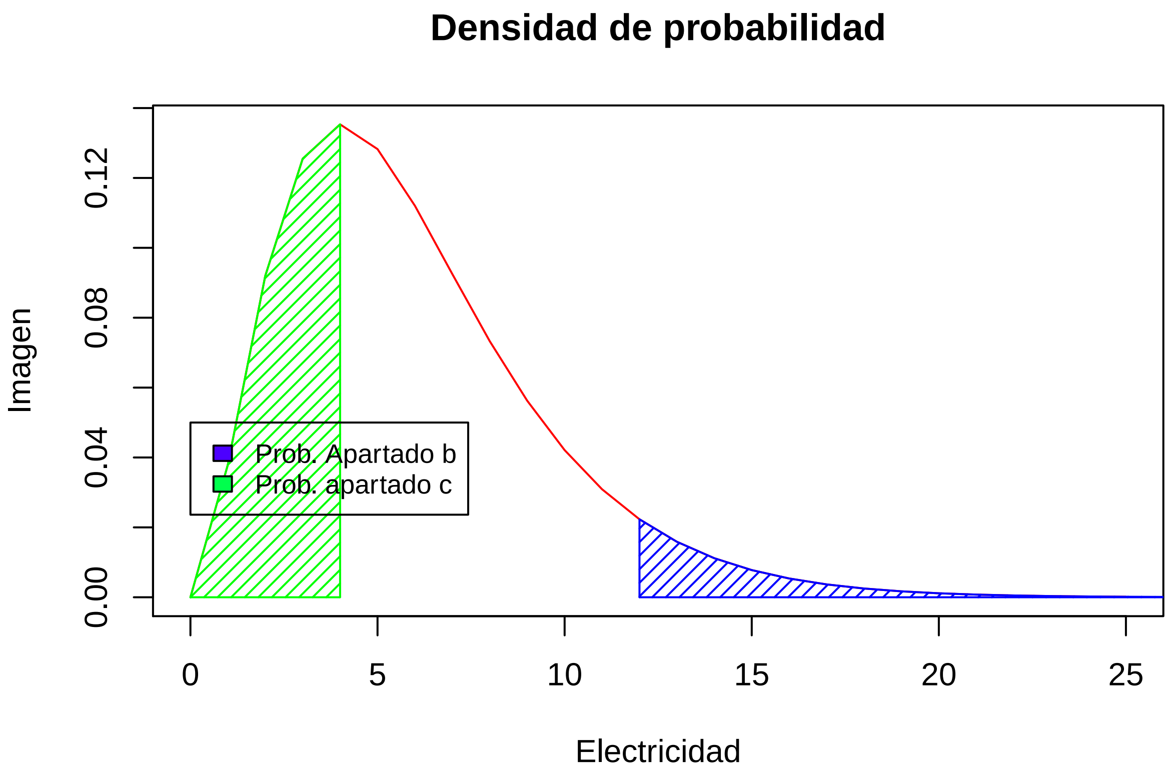
Por último mostraremos las distribuciones haciendo uso de las funcionalidades de R

```
z <- c(12, seq(12,500), 500)
p <- c(0, dgamma(seq(12,500), shape=a, scale=b),0)

plot(0:25, dgamma(0:25, shape=a, ,scale=b), type='l', col="red", xlab="Electricidad", ylab="Imagen", ma
polygon(z, p, density=c(20, 20), col="blue")

z <- c(0, seq(0,4), 4)
p <- c(0, dgamma(seq(0,4), shape=a, scale=b),0)
polygon(z, p, density=c(20, 20), col="green")

legend(0, 0.05, legend=c("Prob. Apartado b", "Prob. apartado c"),
       col=c("red", "blue"), fill=topo.colors(3), cex=0.8)
```



Cuestiones lectura 6

Lectura 6

Cuestión 1:

a)

```
agricultores <- seq(0:8); agricultores
```

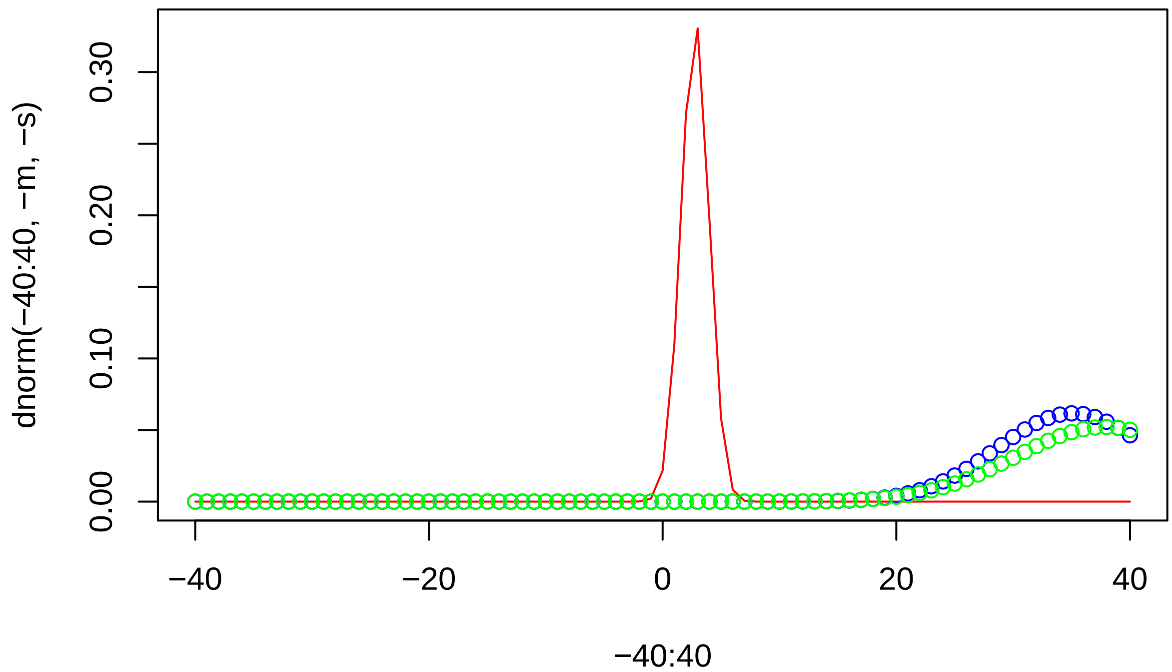
```
## [1] 1 2 3 4 5 6 7 8 9
```

```
c1 <- c(38,23,35,41,44,29,37,31,38)
c2 <- c(45,25,31,38,50,33,36,40,43)
data <- data.frame(agricultores = agricultores, c1 = c1, c2 = c2)

m <- mean(c1) - mean(c2);m
```

```
## [1] -2.777778
```

```
s <- sd(c1) - sd(c2)
plot(-40:40, dnorm(-40:40, -m, -s), type='l', col='red')
points(-40:40, dnorm(-40:40, mean=mean(c1), sd=sd(c1)), type="b", col='blue')
points(-40:40, dnorm(-40:40, mean=mean(c2), sd=sd(c2)), type="b", col='green')
```



b)

Para calcular los intervalos de confianza al no especificarnos la \$ o^2 \$ podemos imaginarnos dos casos primero que en ambas poblaciones \$ o^2\$ son iguales o bien son diferentes.

Estudiaremos a fondo el segundo caso puesto que es la peor situación y despúes analizaremos los otros dos. A su vez veremos reflejada la importanacia de que las muestras sean pareadas.

Primero calcularemos el intervalo no considerando que las muestras son pareadas y que las \$ o^2 \$ son diferentes.

```
x1 <- mean(data$c1); x1
```

```
## [1] 35.11111
```

```
x2 <- mean(data$c2); x2
```

```
## [1] 37.88889
```

```
sd1 <- sd(data$c1)^2
sd2 <- sd(data$c2)^2
a <- (1- 0.95)/2; a
```

```
## [1] 0.025
```

```
V_R_1<-var(data$c1);
V_R_2<-var(data$c2);

n1<-length(data$c1)
n2<-length(data$c2)
```

```r
GL<-((V_R_1/n1+V_R_2/n2)^2)/((V_R_1/n1)^2/(n1-1)+(V_R_2/n2)^2/(n2-1));
ta <- -qt(a, GL); ta
```

```
## [1] 2.124704
```

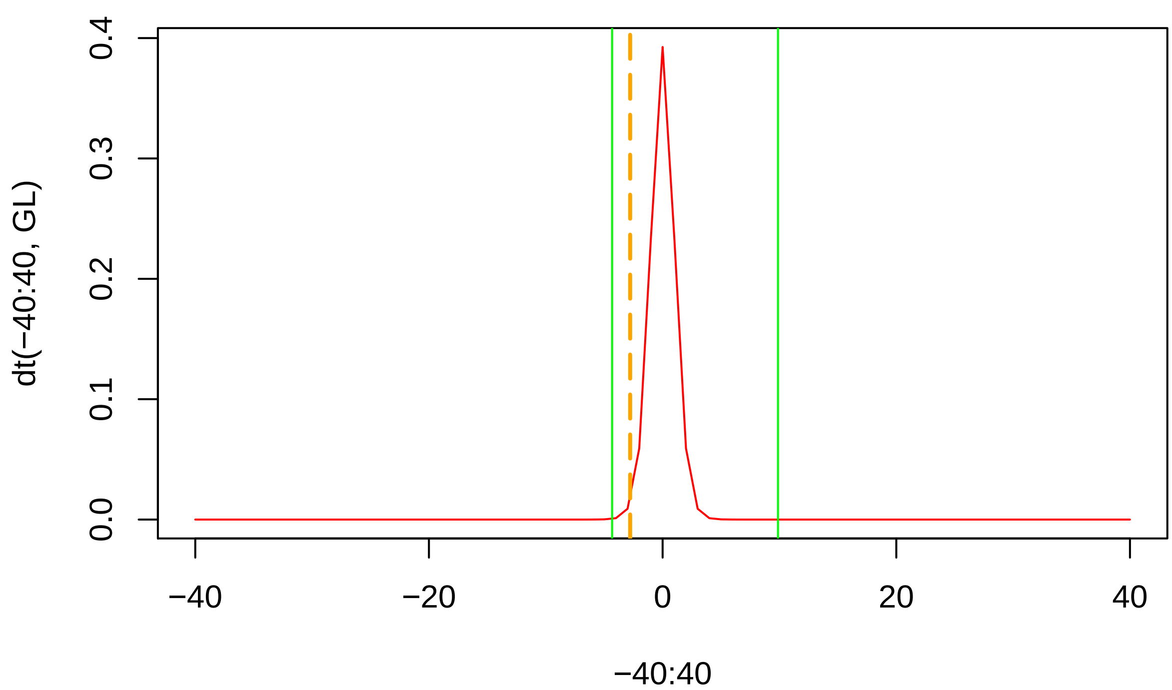
```r
lc1 <- (x2 - x1) - ta*sqrt(sd1/n1 + sd2/n2); lc1
```

```
## [1] -4.321271
```

```r
lc2 <- (x2 - x1) + ta*sqrt(sd1/n1 + sd2/n2); lc2
```

```
## [1] 9.876827
```

```r
plot(-40:40, dt(-40:40, GL), type='l', col='red')
abline(a=-4000, v=lc1, 90, col="green")
abline(a=-4000, v=lc2, 90, col="green")
abline(a=-4000, v=m, 90, col="orange", lty=5, lwd=2)
```



Hemos obtenido el intervalo correspondiente a esta suposicion ahora calculemos lo mismo pero asumiendo que son pareadas (para esto usaremos t.test que nos será mas eficiente que realizar los cáculos de nuevo). Como podemos observar nuestro intervalo decrece considerablemente debido a que al asumir que las muestras son pareadas podemos disminuir los grados de libertad y crear mejores estimaciones.

```r
#realizado antes calculando (muestras no pareadas)
t.test(c1, c2, mu=0, paired=FALSE, var.equal=FALSE, conf.level = 0.95 )
```

```
##
##  Welch Two Sample t-test
```

```
##
## data:  c1 and c2
## t = -0.83137, df = 15.567, p-value = 0.4183
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.876827  4.321271
## sample estimates:
## mean of x mean of y
##  35.11111  37.88889
```

```r
#muestras pareadas
t.test(c1, c2, mu=0, paired=TRUE, var.equal=FALSE, conf.level = 0.95 )
```

```
##
##  Paired t-test
##
## data:  c1 and c2
## t = -1.8209, df = 8, p-value = 0.1061
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6.2955948  0.7400393
## sample estimates:
## mean of the differences
##               -2.777778
```

Podemos repetir lo mismo pero asumiendo que las varianzas poblacionales son iguales.

```r
#sd
t.test(c1, c2, mu=0,  var.equal=TRUE, paired= TRUE, conf.level = 0.95 )
```

```
##
##  Paired t-test
##
## data:  c1 and c2
## t = -1.8209, df = 8, p-value = 0.1061
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6.2955948  0.7400393
## sample estimates:
## mean of the differences
##               -2.777778
```

```r
t.test(c1, c2, mu=0,  var.equal=TRUE, conf.level = 0.95 )
```

```
##
##  Two Sample t-test
##
## data:  c1 and c2
## t = -0.83137, df = 16, p-value = 0.418
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.860794  4.305238
## sample estimates:
## mean of x mean of y
##  35.11111  37.88889
```

c)

Como hemos comentado antes es importante que las muestras sean pareadas pues nos permiten disminuir el grado de libertad cerrando aun mas el intervalo que calculamos.Cuando hacemos este tipo de medidas a su vez lo ideal es que las variables esten interelacionadas.

Cuestión 2:

En el enunciado nos piden que calculemos unos intervalos entre los cuales oscilara $ o^2 $ de la población con una confianza del 98% dándonos los datos de la media muestral, el tamaño de la muestra y la varianza de la misma (puntualizar: nos dan la varianza de la muestra no de la población). Para calcular este intervalo usaremos una distribucion chisq y la siguiente formula $(n-1)*s^2/X^2_{a/2} < o^2 < (n-1)*s^2/X^2_{1-a/2}$

```
x <- 7.25;x
```

```
## [1] 7.25
```

```
a <- 1-0.98;a
```

```
## [1] 0.02
```

```
n <- 25; n
```

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

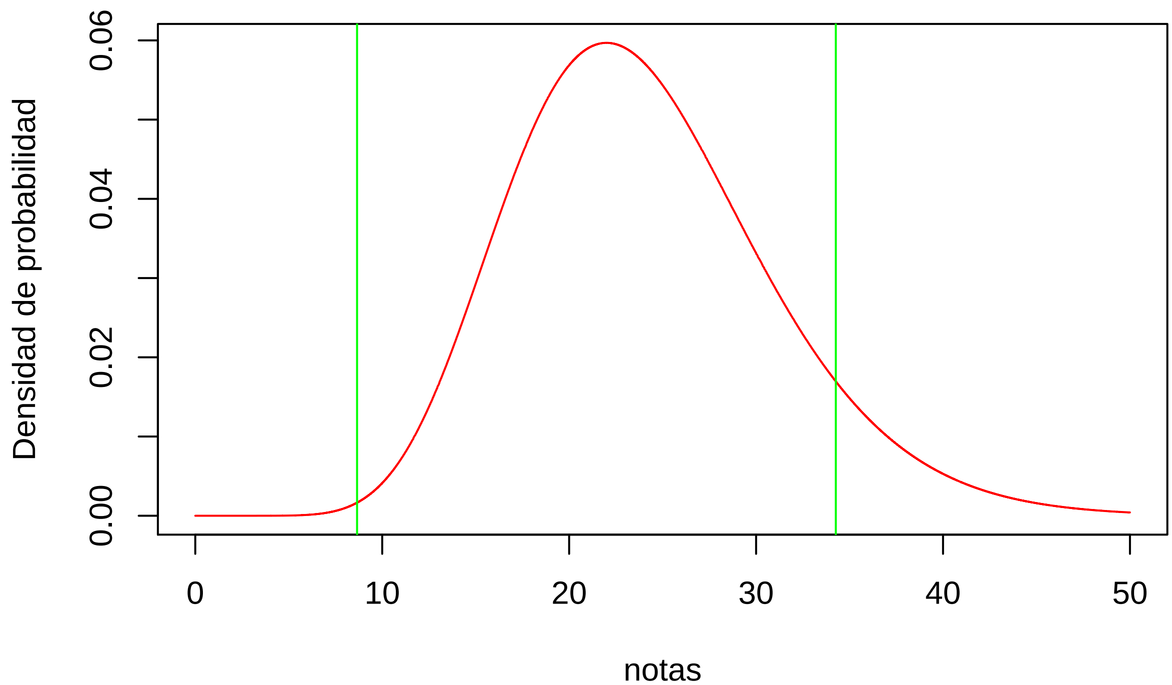
```
lc1<- ((n-1)*15.5)/qchisq(1-a/2, n-1); lc1
```

```
## [1] 8.655225
```

```
lc2<- ((n-1)*15.5)/qchisq(a/2, n-1); lc2
```

```
## [1] 34.26562
```

```
plot(seq(0,50, 0.001), dchisq(seq(0,50, 0.001), n-1), type="l", col="red", xlab="notas", ylab="Densidad
abline(a=1000, v=lc1, 90, col="green")
abline(a=1000, v=lc2, 90, col="green")
```



Cuestión 3:

En esta cuestión nos estan pidiendo una estimación de la proporción verdadera de empresas interesadas en nuestro producto, es decir, queremos saber estimar el porcentaje con un 95% de confianza. Para ello nos están facilitando una muestra de 30 empresas de las cuales 8 se mostraron interesadas. La información que nos solicitan viene dada por la siguiente expresión $p - z_{a/2} * sqrt(p * q/n) < p < p + z_{a/2} * sqrt(p * q/n)$, siendo p la relacion entre las empresas que se han mostrado interesadas y las que no, y q siendo p negada.

```
p <- 8/30; p
```

```
## [1] 0.2666667
```

```
q <- 1-p; q
```

```
## [1] 0.7333333
```

```
a<- (1-0.95)/2; a
```

```
## [1] 0.025
```

```
a<- -qnorm(a);a
```

```
## [1] 1.959964
```

```
lc1 <- p - a*sqrt(p*q/30); lc1
```

```
## [1] 0.1084244
```

```
lc2 <- p + a*sqrt(p*q/30); lc2
```

```
## [1] 0.424909
```

```r
plot(seq(-10,10, 0.001), dnorm(seq(-10,10, 0.001)), type="l", col="red" )
abline(a=1000, v=lc1, 90, col="green")
abline(a=1000, v=lc2, 90, col="green")
```