Estadística Básica con R y R–Commander

(Versión Febrero 2008)

Autores:
A. J. Arriaza Gómez
F. Fernández Palacín
M. A. López Sánchez
M. Muñoz Márquez
S. Pérez Plaza
A. Sánchez Navas

UCA Universidad de Cádiz
Servicio de Publicaciones

Edita: Servicio de Publicaciones de la Universidad de Cádiz
C/ Dr. Marañón, 3
11002 Cádiz

http://www.uca.es/publicaciones

ISBN:

Depósito legal:

Estadística Básica con R y R-commander
(Versión Febrero 2008)
Autores: A. J. Arriaza Gómez, F. Fernández Palacín, M. A. López Sánchez, M. Muñoz Márquez, S. Pérez Plaza, A. Sánchez Navas

http://knuth.uca.es/ebrcmdr

Capítulo 4

Distribuciones de Probabilidad

La existencia de fenómenos o experimentos no determinísticos, donde el conocimiento de las condiciones en las que éstos se desarrollan no determinan los resultados, hace imprescindible el uso de una función que asigne niveles de certidumbre a cada uno de los desenlaces del fenómeno y ahí es donde aparece la *teoría de la probabilidad*. Los experimentos o fenómenos que poseen la característica anterior se denominan aleatorios. Intuitivamente, la concreción numérica del fenómeno mediante la asignación de valores con un cierto criterio, da origen a la *variable aleatoria*. Una correcta proyección de estos conceptos es lo que va a permitir estudiar grandes colectivos a partir de pequeñas partes de ellos, llamadas muestras, dando lugar a lo que se conoce como *inferencia estadística*.

La teoría de la probabilidad y la variable aleatoria van a permitir establecer un amplio catálogo de modelos teóricos, tanto discretos como continuos, a los cuales se van a poder asimilar muchas de las situaciones de la vida real. El estudio de los modelos teóricos, incluyendo la caracterización a través de sus parámetros, el cálculo de probabilidades en sus distintos formatos y la generación de números aleatorios, van a facilitar enormemente el análisis de estas situaciones reales. Ese será el objetivo del capítulo.

Antes de entrar en materia se describirán una serie de fenómenos que se podrán asimilar a las distribuciones de probabilidad que se describirán en este capítulo.

Ejemplo 4.1

- *Si se contesta al azar un examen tipo test de* 10 *preguntas, donde cada una de ellas tiene* 4 *posibilidades siendo sólo una de ellas cierta, ¿qué número de aciertos es más probable?*

- *Cuando alguien pregunta por el número que salió en el sorteo de la ONCE, la respuesta suele ser la unidad de dicho número: el 7, el 5,...¿cómo se distribuyen las unidades de los premios en el sorteo de la ONCE?*

- *En las oposiciones es frecuente que se realice un sorteo público extrayendo una serie de bolas o papeletas de una urna o bolsa. Imagínese un opositor que se ha preparado* 60 *temas de* 100*, de los que se seleccionan al azar dos de ellos, ¿qué probabilidad tiene el opositor de que sea elegido al menos uno de los temas que lleva preparado?*

- *Sabemos que el servicio de autobuses entre Cádiz y San Fernando tiene salidas cada media hora entre las* 6 *am y las* 12 *pm, una persona que se ha olvidado el reloj en casa llega a la estación de autobuses en Cádiz ¿cuál es la probabilidad de que espere menos de* 10 *minutos para coger el autobús?*

- *Se sabe que las bombillas de bajo consumo de* 14 *w tienen una vida media útil de* 10000 *horas, mientras que las bombillas clásicas por incandescencia de* 60 *w tienen una vida media útil de 1000 horas. Si cada día se encienden unas* 4 *horas ¿cuál es la probabilidad de que después de un año estén funcionando las dos?, ¿y ninguna de ellas?, ¿y al menos una de ellas?, ¿y como mucho una de ellas?*

- *Si se controlan el peso, la edad, la estatura, la talla de pantalón, las horas de estudio, la nota de selectividad, ... de los 350 alumnos que están matriculados en 1º de Empresariales y Económicas en el campus de Cádiz y Jerez, ¿qué estructura tiene su distribución?*

Cada una de las situaciones anteriores conlleva la realización de un experimento aleatorio: “elegir una de las cuatro posibles respuestas en cada una de las preguntas”, “extraer la bola del número de las unidades entre las 10 posibles”, “sacar 2 temas entre 100”, ..., que proporcionan resultados de distinta naturaleza. Así, el número de aciertos que se puede obtener al responder las 10 preguntas “variará” entre 0 y 10, o sea, tiene un número finito de posibles valores, mientras que el tiempo de espera para coger el autobús puede tomar infinitos valores dentro del intervalo $(0, 30)$, sólo condicionado por la precisión de los aparatos de medición. Esto lleva a una primera gran clasificación entre modelos de probabilidad discretos y continuos. El primer problema a resolver será la elección del modelo teórico apropiado para cada caso en estudio.

Para tener un buen manejo matemático de las distintas situaciones que se puedan plantear dada la distinta naturaleza y la diversidad de los resultados que proporcionan los experimentos, se necesita realizar una abstracción cuantificada del experimento. Para ello se asignará a cada uno de los posibles resultados del experimento aleatorio (suceso elemental) un número real. A esta aplicación se le llamará *variable aleatoria* y se designará por X, $X : \Omega \to R$. Así en el primer ejemplo, la variable aleatoria consistiría en asignar al suceso “responder correctamente siete preguntas” el número 7. Esta asignación no es única, se le podría haber asignado otro número, por ejemplo 17, lo que proporcionaría otra variable aleatoria, pero en este caso los valores no serían fácilmente identificables en términos del experimento de partida. Como norma, se intentará que la asignación se realice de la forma más natural posible.

Además, por abuso de lenguaje, se tiende a confundir la aplicación X con los valores del conjunto imagen y se traslada la probabilidad de ocurrencia de un suceso al valor correspondiente de la variable aleatoria; por lo tanto, se puede hablar de la probabilidad de que la variable aleatoria tome un determinado valor. Las probabilidades asociadas a cada uno de los valores de la variable aleatoria pueden ser organizadas como una distribución de probabilidad, expresándose mediante una tabla, una gráfica o una fórmula, denominándose en este último caso, a la regla de correspondencia valores–probabilidades, *función de probabilidad*.

DISCRETAS		
Distribución	**Parámetros**	**En Rcmdr**
Binomial	$n = size$; $p = prob$	binom
Binomial negativa	$n = size$; $p = prob$	nbinom
Geométrica	$p = prob$	geom
Hipergeométrica	$(N, K, n) = (m, n, k)$	hyper
Poisson	$\lambda = lambda$	pois

Tabla 4.1: Tabla de distribuciones discretas

Como se ha indicado, según la naturaleza de la variable aleatoria pueden considerarse distribuciones de probabilidad *discretas* o *continuas*. Las principales distribuciones de probabilidad de variables discretas son: *Binomial*, *Binomial Negativa*, *Geométrica*, *Hipergeométrica* y de *Poisson*. Entre los modelos de variable continua destacan las distribuciones: *Normal*, *T-Student*, *Chi-Cuadrado*, *F-Snedecor*, *Exponencial*, *Uniforme*, *Beta*, *Cauchy*, *Logística*, *Lognormal*, *Gamma*, *Weibull* y *Gumbel*. Todas estas distribuciones están recogidas en **Rcmdr**. Se puede acceder a ellas en: `Distribuciones`→`Distribuciones continuas`, o en `Distribuciones`→`Distribuciones discretas`, o también escribiendo directamente en la ventana de instrucciones el nombre de la distribución, poniendo delante una `d`, si se quiere *la función de densidad*, una `p` para la *función de distribución*, una `q` para los *cuantiles* y una `r` para generar una *muestra aleatoria* de la distribución; además, por supuesto, de los argumentos necesarios en cada caso.

1. Distribuciones discretas

En la tabla 4.1 están resumidas todas las distribuciones contenidas en la versión actual de **Rcmdr**, sus parámetros (el nombre teórico y el usado en el programa) y las instrucciones correspondientes. Para cada una de las distribuciones discretas están disponibles las siguientes opciones:

- **Cuantiles:** Permite calcular el valor de la variable que deja a derecha o a izquierda (según se seleccione) una determinada probabilidad.

- **Probabilidades:** Determina la probabilidad de que la variable tome un valor dado.

- **Gráfica de la distribución:** Genera la gráfica de la función de cuantía o de distribución.

- **Muestra de la distribución:** Genera muestras aleatorias extraídas de la distribución.

- **Probabilidades Acumuladas:** Calcula bien el valor de $P(X \leq x)$ (cola de la izquierda), o bien, $P(X > x)$ (cola de la derecha) para cada valor x.

Con el fin de familiarse con las distribuciones y su uso desde **Rcmdr**, se verán ahora algunos ejemplos representativos de las distribuciones más usuales.

1.1. Distribución Binomial

Ejemplo 4.2

Si un estudiante responde al azar a un examen de 8 preguntas de verdadero o falso.

a) ¿Cuál es la probabilidad de que acierte 4?

La variable X="número de aciertos" sigue una distribución Binomial de parámetros $n = 8$ y $p = 1/2$. Para calcular las probabilidades en ***Rcmdr*** *se selecciona:* `Distribuciones→Distribuciones discretas→ Distribución binomial→Probabilidades binomiales...`

En este caso se introduce `Ensayos binomiales= 8` *y* `Probabilidad de éxito= 0.5` *y se puede ver que $P(X = 4) = 0{,}2734375$.*

```
>.Table <- data.frame(Pr=dbinom(0:8, size= 8, prob= 0.5))
>rownames(.Table) <- 0:8
>.Table


   Pr
0  0.00390625
1  0.03125000
2  0.10937500
3  0.21875000
4  0.27343750
5  0.21875000
6  0.10937500
7  0.03125000
8  0.00390625
```

b) ¿Cuál es la probabilidad de que acierte 2 o menos?
Se calculan ahora las probabilidades acumuladas: `Distribuciones→ Distribuciones discretas→Distribución binomial→ Probabilidades binomiales acumuladas...` *Para calcular la probabilidad de que acierte 2 preguntas o menos, en la ventana que aparece, se debe indicar* `Valor de la variable= 2` *y* `Ensayos binomiales= 8`*, dejando marcada la opción* `Cola izquierda`*.*

```
>pbinom(c(2), size= 8, prob= 0.5, lower.tail=TRUE)
[1] 0.1445313
```

c) ¿Cuál es la probabilidad de que acierte 5 o más?
Para determinar la probabilidad de que acierte 5 o más preguntas se realiza el mismo procedimiento, pero señalando en la ventana emergente `Valor de la variable= 4`*, y* `Ensayos binomiales= 8`*, tomándose la opción* `Cola Derecha`*.*

```
>pbinom(c(4), size=8, prob=0.5, lower.tail=FALSE)
[1] 0.3632813
```

1.2. Distribución de Poisson

Ejemplo 4.3

Una cierta área de Estados Unidos es afectada, en promedio, por 6 hura-

canes al año. Encuentre la probabilidad de que en un determinado año esta área sea afectada por:

a) Menos de 4 huracanes.

Se define la variable $X =$"número de huracanes por año" y se sabe que ésta se distribuye mediante una Poisson, porque describe el número de éxitos por unidad de tiempo y porque son independientes del tiempo desde el último evento. Se calcularán ahora las probabilidades:
Como en el caso anterior se señala `Probabilidades binomiales acumuladas...` *tomando ahora en la ventana emergente* `Valor(es) de la variable=` *4, y* `Media=` *6, para la opción* `Cola izquierda`*.*

```
>ppois(c(3), lambda = 6, lower.tail=TRUE)
[1] 0.1512039
```

b) Entre 6 y 8 huracanes.

Para calcular la probabilidad de que ocurran entre 6 y 8 huracanes, se pueden sumar las probabilidades $P(X = 6) + P(X = 7) + P(X = 8)$ o restar las probabilidades acumuladas, con la opción `Cola izquierda`*, $P(X \leq 8) - P(X \leq 5)$. Como antes se realizan en primer lugar las probabilidades acumuladas y se restan los resultados obtenidos:*

```
>a <- ppois(c(8), lambda = 6, lower.tail=TRUE)
>b <- ppois(c(5),lambda = 6, lower.tail=TRUE)
>a-b
[1] 0.4015579
```

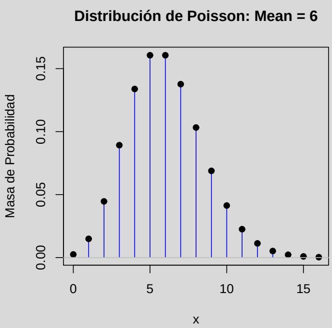

Fig. 4.1: Distribución de Poisson

c) Represente la función de probabilidad de la variable aleatoria que mide el número de huracanes por año. La gráfica se realiza en `Distribuciones→ Distribuciones discretas →Distribución de Poisson→Gráfica de la distribución de Poisson...` *(figura 4.1).*

1.3. Distribución Hipergeométrica

Ejemplo 4.4

En un juego se disponen 15 globos llenos de agua, de los que 4 tienen premio. Los participantes en el juego, con los ojos vendados, golpean los globos con un palo por orden hasta que cada uno consigue romper 2.

a) ¿Cuál es la probabilidad de que el primer participante consiga un premio?

Para el primer participante la variable X="número de premios conseguidos entre 2 posibles" sigue una distribución Hipergeométrica de parámetros $m = 11, n = 4, K = 2$. Para obtener respuesta a las cuestiones en **Rcmdr** *se selecciona:* `Distribuciones→ Distribuciones discretas→Distribución hipergeométrica...`
Para calcular la probabilidad de que consiga un sólo premio se elige la opción `probabilidades hipergeométricas...`*, con* `m(número de bolas blancas en la urna)= 11`*,* `n(número de bolas negras en la urna)= 4` *y* `k(número de extracciones)= 2`*, resultando $P(X = 1) = 0{,}41904762$.*

```
>.Table <- data.frame(Pr=dhyper(0:2, m=11, n=4, k=2))

>rownames(.Table) <- 0:2

>.Table

          Pr
0 0.05714286
1 0.41904762
2 0.52380952
```

b) Construya la gráfica de la función de distribución.

Ésta se obtiene en `Distribuciones→Distribuciones discretas→ Distribución hipergeométrica→Gráfica de la distribución hipergeométrica...`*, marcando la opción* `gráfica de la función de distribución` *(figura 4.2).*

c) Si el primer participante ha conseguido sólo un premio, ¿cuál es la probabilidad de que el segundo participante consiga otro?

Para el segundo participante la variable seguirá una hipergeométrica de parámetros `m= 10`*,* `n= 3` *y* `k= 2`*, resultando $P(X = 1) = 0{,}38461538$.*

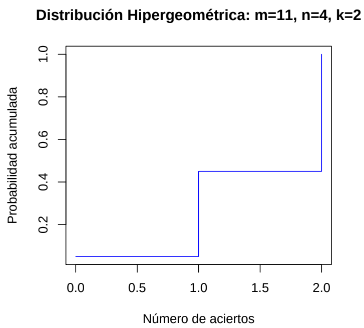

Figura 4.2: Distribución hipergeométrica

1.4. Distribución Geométrica. Distribución Binomial Negativa

Ejemplo 4.5

Un vendedor de alarmas de hogar tiene éxito en una casa de cada diez que visita. Calcula:

a) La probabilidad de que en un día determinado consiga vender la primera alarma en la sexta casa que visita.

Se define la variable X=“número de casas que visita antes de conseguir vender la primera alarma”, que sigue una distribución Geométrica con `Probabilidad de éxito= 0.1`*. Se selecciona en* **Rcmdr** `Distribuciones`→`Distribuciones discretas`→ `Distribución geométrica`→`Probabilidades geométricas...`.

Habrá que calcular la probabilidad de que tenga 5 fracasos antes del primer éxito, obteniendo de la tabla la probabilidad $P(X = 5) = 5{,}904900e{-}02$.

b) La probabilidad de que no venda ninguna después de siete viviendas visitadas.

La variable X=“número de alarmas vendidas en 7 viviendas” sigue una distribución Binomial con `Ensayos binomiales= 8` *y* `Probabilidad de éxito= 0.1`*, luego en nuestro caso se tiene* $P(X = 0) = 0{,}4782969$.

c) Si se plantea vender tres alarmas, ¿cuál es la probabilidad de que consiga su objetivo en la octava vivienda que visita?

CONTINUAS		
Distribución	**Parámetros**	**En Rcmdr**
Normal	$\mu = mean$; $\sigma = sd$	norm
T-Student	$n = df$	t
Chi-Cuadrado	$n = df$	chisq
F-Snedecor	$n = df1$; $m = df2$	f
Exponencial	$\lambda = rate$	exp
Uniforme	$(a, b) = (min, max)$	unif
Beta	$p = shape1$; $q = shape2$	beta
Cauchy	$t = location; s = scale$	cauchy
Logística	$t = location; s = scale$	logis
Lognormal	$\mu = meanlog;\sigma = sdlog$	lnorm
Gamma	$p = shape$; $\alpha = scale$	gamma
Weibull	$p = shape$; $\alpha = scale$	weibull
Gumbel	$p = shape$; $\alpha = scale$	gumbel

Tabla 4.2: Tabla de distribuciones continuas

Para abordar esta cuestión, se define la variable Y= "número de casas que visita antes de conseguir vender la tercera alarma". Esta variable sigue una distribución Binomial Negativa de parámetros `Número de éxitos= 3`*,* `Probabilidad de éxito= 0.1`*. En* **Rcmdr** *se selecciona* `Distribuciones`→ `Distribuciones discretas`→`Distribución binomial negativa`→ `Probabilidades binomiales negativas...`*, de donde:* $P(Y = 5) = 1{,}240029e{-}02$.

2. Distribuciones continuas

En la tabla 4.2 están resumidas todas las distribuciones continuas contenidas en la versión actual de **Rcmdr**, sus parámetros (el nombre teórico y el usado en el programa) y las correspondientes instrucciones.

Para cada una de las distribuciones continuas están disponibles las siguientes opciones:

- **Cuantiles:** Permite calcular el valor de la variable que deja a derecha o a izquierda (según seleccionemos) una determinada probabilidad.
- **Probabilidades:** Determina la probabilidad que queda acumulada a izquierda (o a derecha) de un valor dado.
- **Gráfica de la distribución:** Genera la gráfica de la función de densidad o de distribución.
- **Muestra de la distribución:** Genera muestras aleatorias extraídas de la distribución.

2.1. Distribución Normal

Trabajando directamente en **R**, para calcular los cuantiles normales se usaría `qnorm`, agregando a ésta los argumentos necesarios. En concreto, para hallar el valor que, en una $N(0,1)$, deja en la cola izquierda una probabilidad de 0,25:

```
qnorm(c(.25), mean = 0, sd = 1, lower.tail = TRUE)
```

R–Nota 4.1

lower.tail = TRUE usa la cola de la izquierda, mientras que lower.tail = FALSE usa la derecha. Los parámetros lower.tail = TRUE, mean = 0 y sd = 1 pueden ser omitidos, pues son los valores por defecto en esta función.

Ejemplo 4.6

Una empresa está buscando personal para su departamento de marketing. El perfil solicitado es el de sujetos extrovertidos y creativos. Se han presentado 50 candidatos y la empresa ha establecido como criterio *de selección el que los candidatos superen el percentil 80 en creatividad y extroversión. Sabiendo que la variable extroversión (X) se distribuye según una Normal de media 5 y desviación típica 1, que la variable creatividad (Y) sigue una t-Student de 10 grados de libertad y que las puntuaciones de creatividad y extroversión son independientes:*

a) ¿Cuántos candidatos serán seleccionados?

Al ser X e Y independientes, la probabilidad $P(X \geq P_{80} \cap Y \geq P_{80}) = P(X \geq P_{80}) \cdot P(Y \geq P_{80}) = 0{,}20 \cdot 0{,}20 = 0{,}04$. *Como se han presentado* 50 *aspirantes, serán seleccionadas* $0{,}04 \cdot 50 = 2$ *personas.*

b) ¿Qué puntuaciones debe superar un aspirante en creatividad y extroversión para ser admitido?

Según el criterio de selección se debe superar el percentil 80*, en ambas variables, para ser admitido. Se calculará pues el percentil P_{80} de la variable X e Y, utilizando los cuantiles normales para la variable X:*

```
> qnorm(c(.8), mean=5, sd=1, lower.tail=TRUE)
[1] 5.841621
```

y los t-cuantiles para la variable Y:

```
> qt(c(.8), df=10, lower.tail=TRUE)
[1] 0.8790578
```

c) Si se extraen al azar 16 *candidatos, ¿cuál es la probabilidad de que su media aritmética en extroversión sea mayor que* 4,5*?*

Se sabe que al extraer una muestra de una población normal de tamaño n, la media de la muestra, $\bar{X}$, sigue otra distribución normal de media igual que la poblacional y desviación típica $\frac{\sigma}{\sqrt{n}}$. Por lo que en este caso $\bar{X} \sim N(5, \frac{1}{4})$. Como se desea calcular $P(\bar{X} \geq 4{,}5)$, se selecciona `Cola derecha` *en la entrada de* `Probabilidades normales...`

```
> pnorm(c(4.5),mean=5,sd=0.25,lower.tail=FALSE)
[1] 0.9772499
```

d) Dibuje las gráficas de densidad de las variables `Extroversión` *y* `Creatividad`*.*

Para ello se selecciona la función de densidad de ambas variables en `Distribuciones`→`Distribuciones Continuas...`*, obteniéndose las figuras 4.3 y 4.4.*

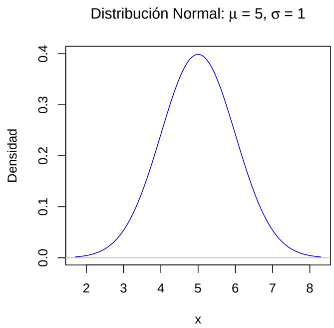

Figura 4.3: Función de densidad de la variable extroversión (normal)

2.2. Distribución Uniforme Continua

Ejemplo 4.7

Una persona informal hace esperar a su pareja aleatoriamente entre 0 *y* 90 *minutos. Harto de esta situación, la persona que sufre la espera se plantea un ultimátum; si al día siguiente su pareja tarda menos de* 15 *minutos mantiene la relación, si la espera está entre* 15 *y* 55 *minutos, decide en la siguiente cita con los mismos criterios, mientras que si tarda más de* 55 *minutos la relación termina en ese momento.*

a) Represente gráficamente la función de densidad de la variable que modeliza esta situación.

Se define la variable X="tiempo de espera", que sigue una distribución uniforme continua definida en el intervalo $(0, 90)$. *En* ***Rcmdr*** *se selecciona* `Distribuciones→Distribuciones continuas→ Distribución uniforme...` *Se elige* ***`Gráfica de la distribución uniforme...`***, *marcando* **`Función de densidad`** *(figura 4.5).*

b) Calcule la probabilidad de que la relación continúe hasta la siguiente cita.

En `Probabilidades uniformes...` *se indica el valor de la variable y los límites del intervalo, dejando la opción* ***`Cola Izquierda`***.

```
> punif(c(55), min=0, max=90, lower.tail=TRUE)
[1] 0.6111111
```

c) Calcule la probabilidad de que la relación termine en la segunda

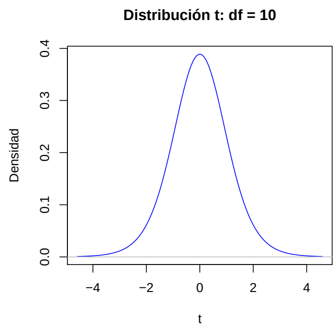

Figura 4.4: Función de densidad de la variable creatividad (t-student)

cita.

b) En `Probabilidades uniformes...` *se indica el valor de la variable y los límites del intervalo, dejando la opción* `Cola Izquierda`*.*

```
> punif(c(55), min=0, max=90, lower.tail=TRUE)
[1] 0.6111111
```

c) Suponiendo que el tiempo de espera en una cita es independiente respecto de otras citas, se calcula la probabilidad $P(15 < X < 55) = P(X < 55) - P(X \leq 15) = 0{,}6111 - 0{,}1666 = 0{,}4445$*, que es la probabilidad de que aplace la decisión para la segunda cita y, en la segunda cita, la probabilidad de que lo deje definitivamente es* $P(X > 55) = 0{,}3888$*, luego multiplicando ambas probabilidades se obtiene el valor pedido* $0{,}1728$*.*

2.3. Distribución Exponencial

Ejemplo 4.8

La duración media de un modelo de marcapasos es de 7 *años.*

a) ¿Cuál es la probabilidad de que dure al menos 5 *años? ¿y menos de* 3*?*

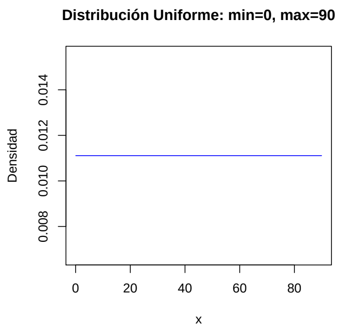

Figura 4.5: Función de densidad

La variable X="tiempo de funcionamiento del marcapasos" sigue una distribución exponencial con parámetro $\lambda = 1/7$. Utilizando la opción `Distribuciones→Distribuciones continuas→ Distribución exponencial→Probabilidades exponenciales...` *se obtiene $P(X \geq 5)$*

```
> pexp(c(5), rate=0.1428, lower.tail=FALSE)
[1] 0.4896815
```

y de igual forma $P(X < 3)$:

```
> pexp(c(3), rate=0.1428, lower.tail=TRUE)
[1] 0.3484493
```

b) Si han transcurrido ya 4 años desde su implantación, ¿cuál es la probabilidad de que dure otros 4?
Teniendo en cuenta que $1-F(x) = e^{-\lambda \cdot x}$, se tiene que $1-F(8) = e^{-8\cdot\lambda} = (e^{-4\cdot\lambda})^2 = (1-F(4))^2$, con lo que $P(X \geq 8/X \geq 4) = (1-F(8))/(1-F(4)) = 1-F(4) = 0{,}5647182$.

c) ¿Cuánto tiempo debería funcionar un marcapasos para estar entre el 10 % de los más duran? Hay que calcular el percentil 90 seleccionando:

`Distribuciones→Distribuciones Continuas→ Distribución exponencial→Cuantiles exponenciales...`, *con las opciones* ***`Probabilidades= 0.9`***, ***`Parámetro de la exponencial= 0.14285`*** *y* ***`Cola Izquierda`***, *o de forma similar,* ***`Probabilidades= 0.1`***, ***`Parámetro de la exponencial= 0.14285`*** *y* ***`Cola Derecha`***,

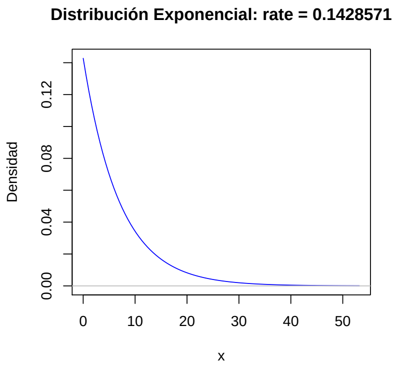

Figura 4.6: Gráfica de la función de densidad de una Exp(0.14285≈ 1/7)

resultando 16,12 *años.*

d) Calcular el valor que deben tener a y b para que $P(X < a) = 0{,}5$ y $P(X > b) = 0{,}32$, De forma análoga al apartado anterior, en el primer caso habría que calcular la mediana, $a = 4{,}852$, y en el segundo, el percentil 68, $b = 7{,}97$. e) Represente la función de densidad de la variable aleatoria asociada. Figura 4.6.

2.4. Distribución t-Student

Ejemplo 4.9

Una variable X sigue una distribución t-Student con 16 grados de libertad.

a) Calcular la mediana y el percentil 85.

Habría que calcular Me de forma que $P(t_{16} \geq Me) = 0{,}5$, para ello se selecciona `Distribuciones→Distribuciones Continuas→ Distribución t→Cuantiles t...`, *con las opciones* `Probabilidades= 0.5`, `Grados de libertad= 16` *y* `Cola Izquierda` *o, de forma similar,* `Probabilidades= 0.5`, `Grados de libertad= 16` *y* `Cola Derecha`, *resulta que el valor de la mediana es* 0.

```
> qt(c(0.5), df=16, lower.tail=TRUE)
[1] 0
```

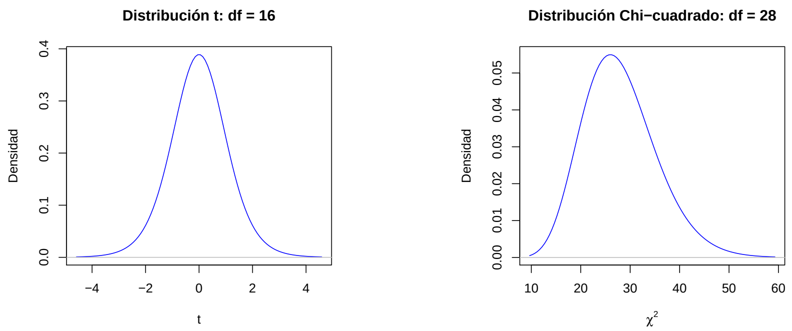

Figura 4.7: Gráfica de la función de densidad t_{16} y χ_{28}

El percentil 85 se calcula de forma parecida:

```
> qt(c(0.85), df=16, lower.tail=TRUE)
[1] 1.071137
```

b) Encontrar el valor de a de forma que $P(-1 < X < a) = 0{,}7$. Para calcular a, se descompone la probabilidad $P(-1 < X < a) = P(X < a) - P(X \leq -1)$, se calcula $P(X \leq -1)$ utilizando la opción `Probabilidades t...`

```
> pt(c(-1), df=16, lower.tail=TRUE)
[1] 0.1660975
```

y, se despeja $P(X < a)$, resultando ser $P(X < a) = 0{,}7 + 0{,}166 = 0{,}866$. Se selecciona ahora la opción `Cuantiles t...`*,*

```
> qt(c(0.866), df=16, lower.tail=TRUE)
[1] 1.147611
```

resultando el valor de `a`$=1{,}147611$.

c) Obtener la gráfica de su función de densidad. ¿Qué similitud tiene con la normal $N(0,1)$?
Como se puede observar en la figura 4.7 su estructura es similar a la $N(0;1)$ con la particularidad de que en la zona central la t_{16} se encuentra por debajo de la normal, consecuencia de tener una varianza mayor.

2.5. Distribución Chi-cuadrado. Distribución F-Snedecor

Ejemplo 4.10

La variable X sigue una distribución Chi-cuadrado con 28 grados de libertad.

a) Calcule la probabilidad de que X sea mayor de 7,5.

La probabilidad pedida $P(\chi_{28} > 7{,}5)$, se obtiene en `Distribuciones`→`Distribuciones Continuas`→`Distribución Chi-cuadrado`→`Probabilidades Chi-cuadrado...`, *con las opciones* `Valor(es) de la variable= 7.5`, `Grados de libertad= 28` *y* `Cola derecha`. *Su valor es* 0,9999611.

```
> pchisq(c(7.5), df=28, lower.tail=FALSE)
[1] 0.9999611
```

b) Obtenga la función de densidad, ¿qué características se observan?. Otra variable Y sigue una distribución F de Snedecor con $n_1 = 8$ y $n_2 = 14$ grados de libertad, si se representa su función de densidad.

Como se puede observar en la figura 4.7 sólo toma valores positivos y es asimétrica con forma campaniforme, salvo para $n \leq 2$. c) ¿Qué similitudes hay entre las gráficas?

Como se aprecia en 4.8, en general, sus características son muy similares a la función de densidad de la χ^2.

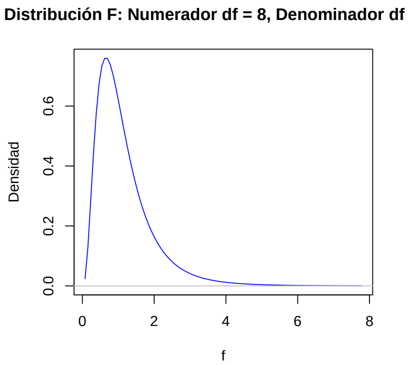

Figura 4.8: Función de densidad $F_{8,14}$

3. Generación de valores aleatorios

Hay situaciones donde es necesario generar valores aleatorios que sigan un determinado patrón y que permitan estudiar el comportamiento de determinados modelos, simular situaciones de laboratorio, generar la distribución de una combinación de variables, comparar valores muestrales con los extraídos de la verdadera población en estudio, ... En **Rcmdr**, para cada una de las distribuciones de probabilidad que tiene implementadas, se puede seleccionar la opción `Muestra de una distribución...`. Así, para generar una muestra de tamaño 15 de una distribución uniforme en el intervalo $[0, 1]$, se selecciona en `Distribuciones`→`Distribuciones continuas`→`Distribución uniforme`→`Muestra de una distribución uniforme...`, y se introducen los parámetros, en este caso, para obtener los datos en formato de columna, `Mínimo= 0`, `Máximo= 1`, `Número de muestras (filas)= 15` y `Número de observaciones (columnas)= 1`.

```
> Muestras_uniformes <- as.data.frame(matrix(runif(15*1,
min=0, max=1), ncol=1))
> rownames(Muestras_uniformes) <- paste(‘‘sample’’, 1:15,
sep=‘‘’’)
> colnames(Muestras_uniformes) <- ‘‘obs’’
```

Para mostrarlos en pantalla se escribe en la ventana de instrucciones el nombre que se le haya asignado a la muestra:

```
> Muestras_uniformes
        obs
sample1 0.22597988
sample2 0.65997127
sample3 0.07038248
sample4 0.52902704
sample5 0.04517561
sample6 0.73990437
sample7 0.90452613
sample8 0.60055627
sample9 0.99432508
sample10 0.70652675
sample11 0.97110556
sample12 0.24558711
sample13 0.68375576
sample14 0.95487024
sample15 0.80651304
```

O también se puede pulsar el botón `Visualizar conjunto de datos` en **Rcmdr**. De la misma forma se podrían generar muestras aleatorias para el resto de las distribuciones de probabilidad.

4. Ejercicios

4.1 Se responde al azar un examen tipo test de 10 preguntas donde en cada una de ellas se plantean 4 posibilidades siendo sólo una de ellas cierta. Si se responden todas las preguntas y, las preguntas con respuestas correcta suman un punto mientras que las contestadas incorrectamente restan un cuarto de punto, se pide:

a) La variable aleatoria asociada.

b) Las gráficas de la función de cuantía y distribución y coméntelas.

c) La probabilidad de obtener 3 aciertos.

d) La probabilidad de aprobar.

e) ¿Qué número de aciertos es más probable?

f) ¿Cuántos aciertos debe tener para quedar por encima de la mitad de la clase?

g) ¿Y por encima de un tercio de la clase?

4.2 Dada la distribución $B(10; 0{,}4)$, calcule las siguientes probabilidades:

a) $P(X \leq 8)$

b) $P(2 < X \leq 5)$

c) $P(X \geq 7)$

4.3 Un conocido fumador gorrón ha explotado tanto a sus compañeros que por término medio cada uno de ellos le da un cigarrillo de cada diez veces que éste les pide.

a) ¿Cuál es la probabilidad de que consiga 1 cigarrillo en menos de 5 intentos?

b) Si pretende hacer acopio de cigarrillos para el fin de semana, ¿cuántas veces, en promedio, tendrá que pedir tabaco para conseguir 20 unidades?

4.4 En las oposiciones es frecuente que se realice un sorteo público extrayendo una serie de bolas o papeletas de una urna o bolsa. Imagínese que un opositor se ha preparado 60 temas entre 100, de los que se seleccionan al azar dos temas. Se pide:

a) La variable aleatoria asociada.

b) Las gráficas de la función de cuantía y distribución y coméntelas.

c) La probabilidad de que le salga uno de los temas que lleva preparado.

d) La probabilidad de que le salgan dos de los temas que lleva preparado.

e) ¿Qué ocurre con la probabilidad anterior si aumenta el número de temas preparados a 80?

4.5 A un establecimiento de apuestas deportivas llega 1 cliente cada 3 minutos por término medio.

a) ¿Cuál es la probabilidad de que en un periodo de 5 minutos lleguen más de 5 clientes?

b) ¿Cuál es el número más probable de llegadas en media hora?

4.6 Las compañías aéreas acostumbran a reservar más plazas de las existentes en sus vuelos, dado el porcentaje de anulaciones que se produce. Si el porcentaje medio de anulaciones es del 5 %, ¿cuántas reservas deberá hacer una compañía para un vuelo con 200 plazas, si quiere con una probabilidad del 97 % que todos sus clientes tengan cabida en dicho vuelo?

4.7 El servicio de reclamaciones de una asociación de consumidores recibe por término medio 3 quejas a la hora.

a) Calcule la probabilidad de que en 1 hora no reciba ninguna reclamación.

b) Calcule la probabilidad de que en 2 horas reciba entre 2 y 6 reclamaciones.

4.8 En una pecera hay 10 peces machos y 8 hembras, si se extraen aleatoriamente 5 peces, calcule la probabilidad de que 3 sean machos y 2 hembras.

4.9 Un jugador apuesta 5€ por tirada a un número de los 37 que componen la ruleta, si acierta, gana 180€. Calcule los beneficios esperados al cabo de 100 jugadas.

4.10 El servicio de autobuses entre Cádiz y San Fernando tiene salidas cada media hora entre las 6 am y las 12 pm, una persona que se ha olvidado el reloj en casa llega a la estación de autobuses en Cádiz, se pide:

a) La variable aleatoria adecuada para esta situación.

b) Las gráficas de la función de densidad y distribución y coméntelas.

c) ¿Cuál es su media? ¿y su mediana? ¿y su moda?

d) La probabilidad de que espere menos de 10 minutos.

e) La probabilidad de que espere más de 15 minutos, pero menos de 20.

f) ¿Cuál es la probabilidad de que espere exactamente 11 minutos y medio?

4.11 Se sabe que las bombillas de bajo consumo de 14 w tienen una vida útil media de 10000 horas, mientras que las bombillas clásicas por incandescencia de 60 w tienen una vida útil media de 1000 horas.Si cada día se encienden unas 4 horas, en esta situación

a) Defina la variable aleatoria asociada.

b) Obtenga las gráficas de la función de densidad y distribución y coméntelas.

c) ¿Cuál es su media? ¿y su mediana?

d) ¿Cuál es la probabilidad de que después de un año estén funcionando?

4.12 ¿Cuál es la probabilidad de que de 10 personas elegidas al azar al menos 2 cumplan años en el mes de Enero?

4.13 Durante la Segunda Guerra Mundial los alemanes bombardearon repetidas veces Londres. Los expertos demostraron que se trataba de bombardeos indiscriminados y que caían en cada acción y por término medio 2 bombas por cada cuadrícula de 100 metros de lado. En vista a lo anterior, calcule la probabilidad de que en una cierta cuadrícula de 50 metros de lado no haya caído ninguna bomba durante un bombardeo.

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

a) $P(2 \leq X \leq 5)$
b) $P(X \geq 3)$
c) $P(X \leq -2)$

4.15 La centralita de un programa de televisión que premia aquellos concursantes que llaman dando la respuesta correcta de un concurso, atiende 1 de cada 10 llamadas que se realizan.

a) ¿Qué número medio de llamadas se tendrán que realizar para ser atendido?

b) ¿Cuál es la probabilidad de ser atendido a la primera?

4.16 Calcule en los siguientes casos el valor de a, sabiendo que $X \sim N(1,5)$.

a) $P(0 \leq X \leq a) = 0{,}28$
b) $P(1 - a \leq X < 1 + a) = 0{,}65$

4.17 Se sabe que la alarma de un reloj saltará en cualquier momento entre las siete y las ocho de la mañana. Si el propietario del reloj se despierta al oír dicha alarma y necesita, como mínimo, veinticinco minutos para arreglarse y llegar al trabajo,

a) ¿Cuál es la probabilidad de que llegue antes de las ocho?

b) Si el dueño del reloj sigue programando el reloj de la misma manera durante 10 días, calcule el número más probable de días en que llegará después de las ocho.

4.18 Si se controlan el peso, la edad, la estatura, talla de pantalón, horas de estudio, nota de selectividad, ... de los 350 alumnos que están matriculados en 1º de Empresariales y Económicas en el campus de Cadiz y Jerez. ¿Qué estructura tiene su distribución?

4.19 De una tribu indígena se sabe que los hombres tienen una estatura que se distribuye según una ley normal con media 1,70 y desviación típica σ. Si a través de estudios realizados se conoce que la probabilidad de que su estatura sea mayor a 1,80 es 0,12, calcule la probabilidad de que un individuo elegido al azar mida entre 1,65 y 1,75.

4.20 Calcule la probabilidad de obtener más de 200 seises en 1200 lanzamientos de un dado no trucado.

4.21 Genere muestras de tamaño 10, 100, 500 y 1000 de una población que sigue una distribución normal de media 3,5 y desviación típica 2. Estudie el comportamiento de la media y desviación típica en las cuatro muestras.

4.22 Obtenga una muestra aleatoria de tamaño 50 para una característica poblacional que sigue una distribución binomial de parámetros $n = 12$ y $p = 0{,}7$. Calcule su media y desviación típica comparándolas con los respectivos valores poblacionales. Además, represente los datos mediante un diagrama de barras y compare los resultados con los observados en la gráfica de la función de cuantía de la distribución binomial. ¿Qué ocurre si se aumenta el tamaño de la muestra a 500?