Práctica de laboratorio nº3

1. *Objetivos*

Usaremos OpenOffice Calc para calcular probabilidades asociadas a las distribuciones discretas y continuas que se han estudiado en la teoría.

2. *Distribuciones discretas*

DISTR.BINOM (x;n;p;0)	Pr [X = x], X → B (n; p)
DISTR.BINOM (x;n;p;1)	Pr [X ≤ x], X → B (n; p)
POISSON (x;λ;0)	Pr [X = x], X → P (λ)
POISSON (x;λ;1)	Pr [X ≤ x], X → P (λ)

En el apartado de funciones estadísticas, Calc dispone de varias funciones que permiten calcular probabilidades asociadas a las distribuciones discretas que se han estudiado en la teoría: la binomial (y su caso particular, la distribución de Bernoulli) y la de Poisson. Para la **distribución binomial** la función de Calc se denomina *DISTR.BINOM*, y para la de Poisson la función de Calc se denomina *POISSON*. Ambas funciones tienen un comportamiento similar y solamente se diferencian en los parámetros que determinan cada distribución.

Para comprender correctamente el modo de trabajar con estas funciones hay que tener en cuenta el significado del parámetro de la función llamado *Acumulado*; puede tomar como argumento el valor 'VERDADERO' (o, equivalentemente, un '1') o 'FALSO' (o un '0'). Si se indica 'FALSO', la probabilidad que nos devolverá el programa es la de que la variable sea exactamente igual al valor considerado, es decir, Pr [X = k]. Por el contrario, si en la casilla *Acumulado* se indica 'VERDADERO', Calc nos devolverá la probabilidad acumulada hasta el valor considerado, es decir, Pr [X ≤ k].

El resto de probabilidades que queramos calcular (mayor que, mayor o igual que, etc.) debemos plantearlas "a mano" a partir de las probabilidades del tipo "igual" y "menor o igual" que nos calcula el programa.

Ejemplo: *Se lanzan 10 monedas. Determina la probabilidad de que el número de caras sea exactamente 6 y la probabilidad de que el número de caras sea mayor o igual que 5.* Evidentemente, llamando X='número de caras obtenidas con las 10 monedas', se nos pide que calculemos Pr [X = 6] y Pr [X ≥ 5]. La distribución de X es una binomial de parámetros n = 10 y p = 0,5. La primera probabilidad, Pr [X = 6], nos la calculará directamente el programa, pues es del tipo “igual que”. Así, bastará con insertar la función *DISTR.BINOM*, como se aprecia en la ilustración 1, con lo que obtendremos el resultado de 0’21.

En cuanto a la segunda probabilidad del ejemplo, Pr [X $\geq$ 5], no es del tipo “igual” ni “menor o igual”, por lo que habrá que transformarla adecuadamente para conseguir expresarla de forma que la podamos calcular con Calc. En este caso, dado que trabajamos con una variable discreta que sólo toma los valores 0, 1, 2, …, 10, se verifica que Pr [X $\geq$ 5] = 1 – Pr [X < 5] = 1 – Pr [X $\leq$ 4]. De esta forma, Pr [X $\leq$ 4] se puede calcular con Calc para (posteriormente o en la misma celda) restárselo a 1. La operación (calculando en un solo paso 1 menos la probabilidad) se observa en la Ilustración 2, consiguiéndose un resultado de 0,38 para Pr [X $\leq$ 4] y de 0,62 para 1 – Pr [X $\leq$ 4].

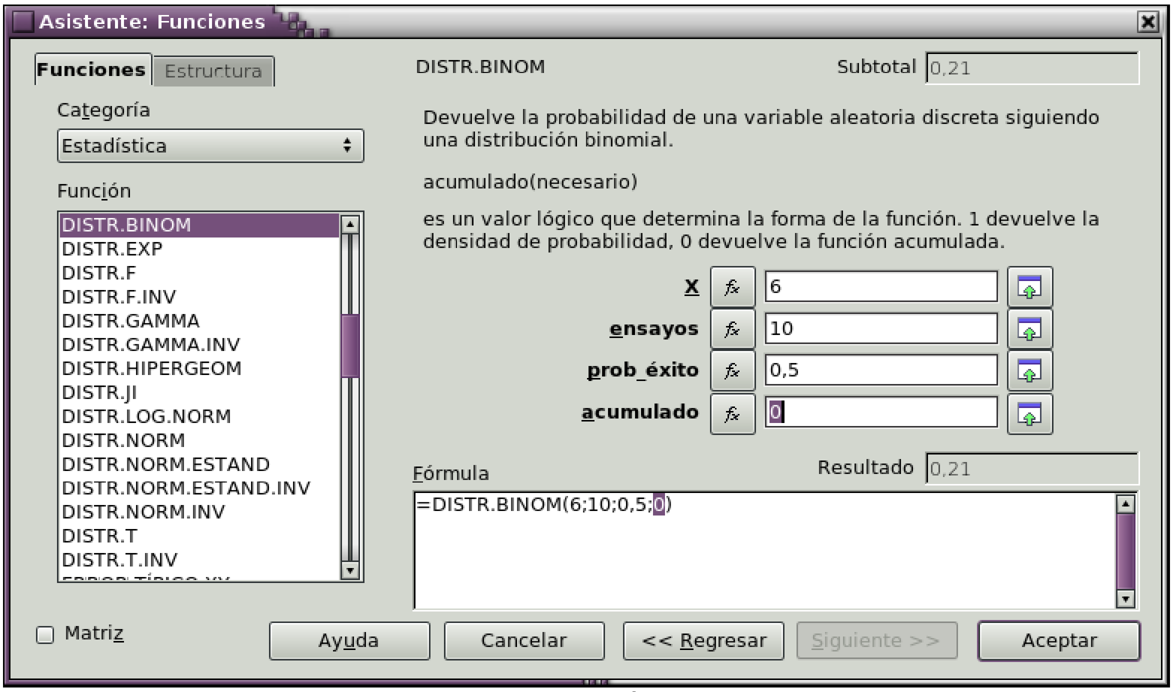

Ilustración 1

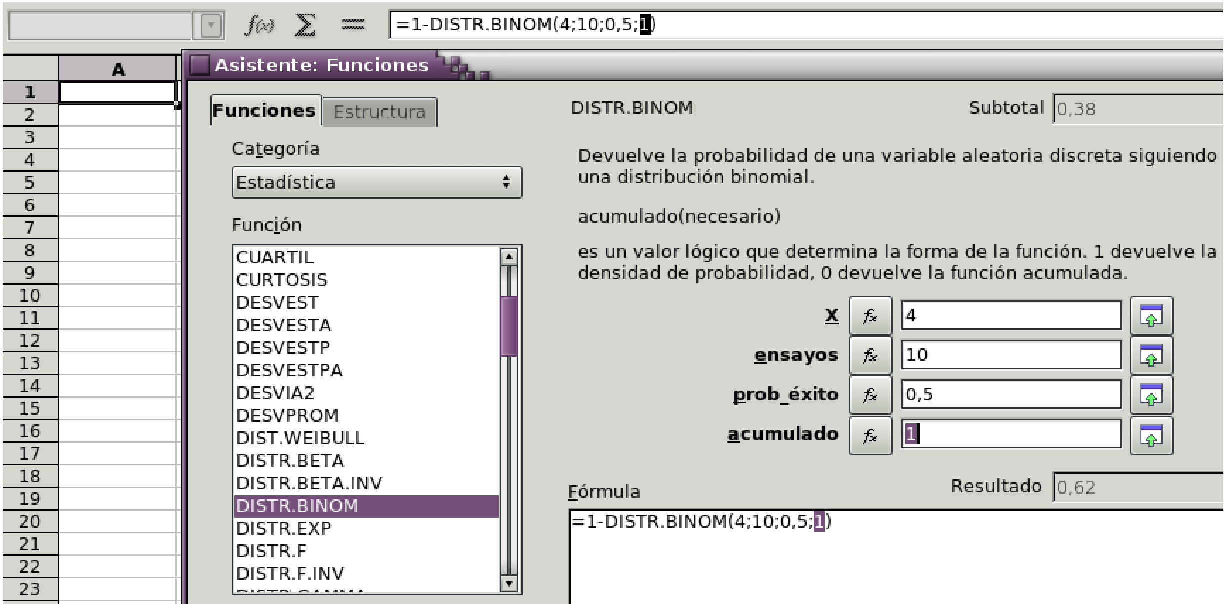

Ilustración 2

Otras probabilidades se obtienen de forma no tan directa.

Ejemplo: *Se lanzan 10 monedas. Determina la probabilidad de que el número de caras obtenidas sea inferior a 6 y la probabilidad de que el número de caras esté entre 7 y 9 (ambos incluidos).* Suponiendo, como en el ejemplo anterior, que el número de caras obtenidas sigue una distribución binomial de parámetros 10 y 0,5, la primera probabilidad se puede escribir como Pr [X<6]. Puesto que Calc no calcula directamente una probabilidad de este tipo, debemos reescribirla en términos que sí sean calculables. Como la variable binomial sólo toma valores enteros (en concreto, esta variable toma los valores 0, 1, 2, ..., 10) es evidente que Pr [X<6] = Pr [X $\leq$ 5] y esta probabilidad sí se puede calcular directamente con *DISTR.BINOM* (utilizando un 1 para la entrada *Acumulado);* se ofrece el resultado en la Ilustración 3.

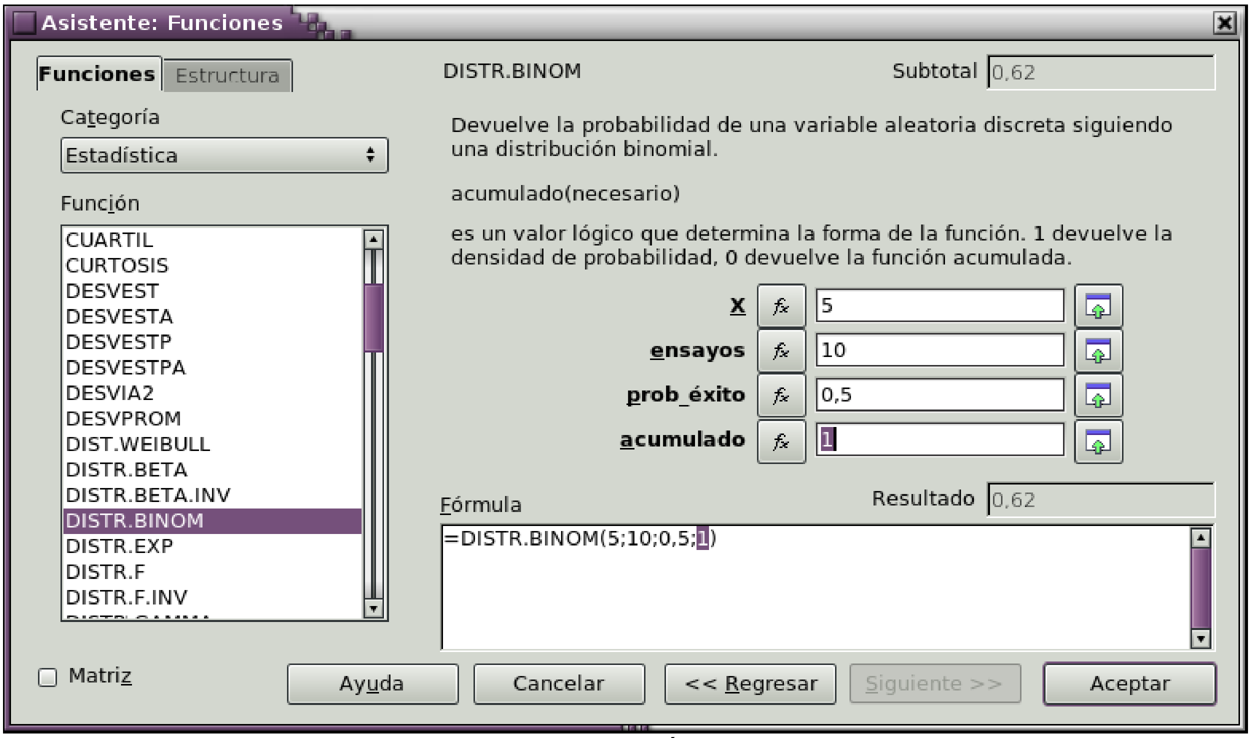

Ilustración 3

La probabilidad de que el número de caras esté entre 7 y 9, ambos inclusive, es Pr [7 ≤ X ≤ 9]. Tampoco se puede calcular directamente con Calc, pero se puede reescribir de varias formas. La primera, como Pr [7 ≤ X ≤ 9] = Pr [X=7] + Pr [X=8] + Pr [X=9], lo que permite calcularla como la suma de tres probabilidades que sí son calculables con la función *DISTR.BINOM* (las tres con un 0 en la entrada *Acumulado*). En la Ilustración 4 observamos el cálculo de la primera de ellas, mientras que en la Ilustración 5 se calcula la suma de las tres probabilidades.

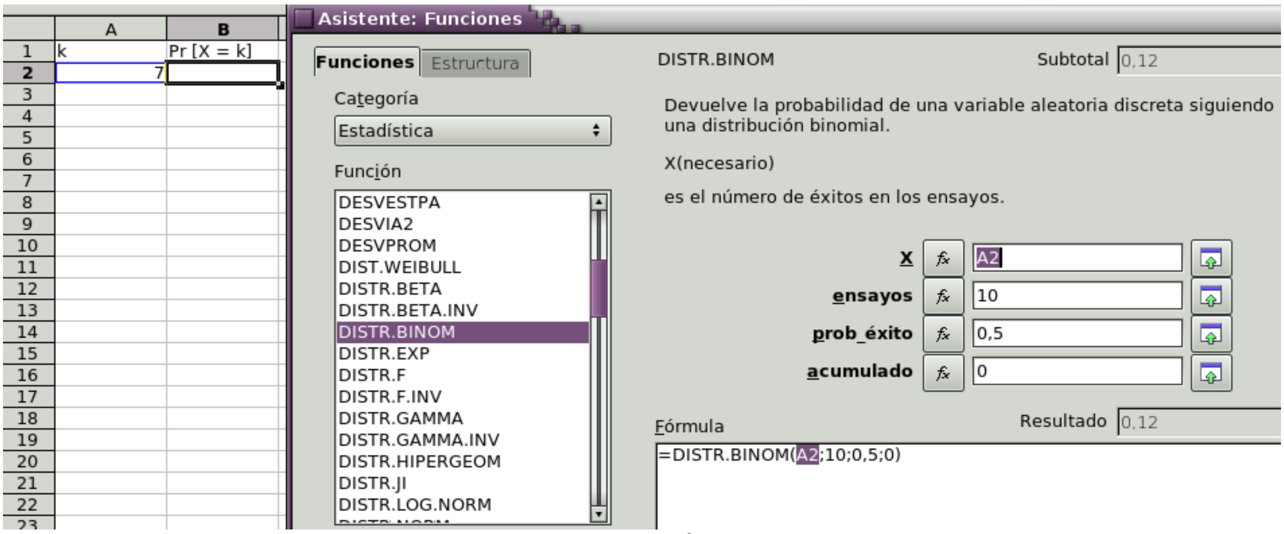

Ilustración 4

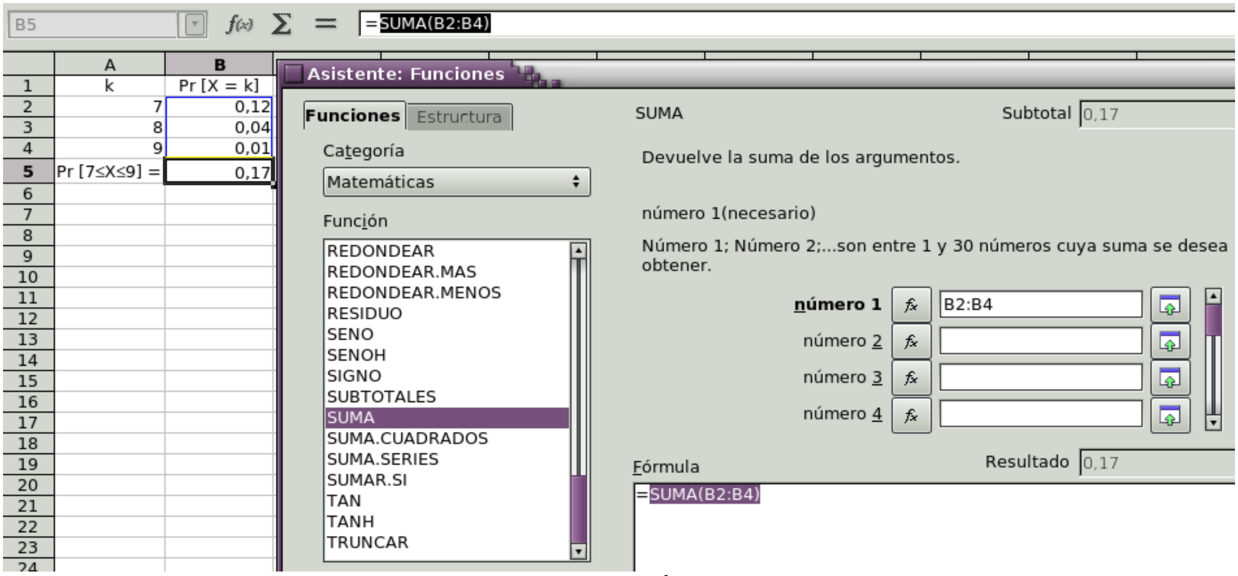

Ilustración 5

Otra posibilidad es escribir Pr [7 $\leq$ X $\leq$ 9] = Pr [X $\leq$ 9] – Pr [X < 7] = Pr [X $\leq$ 9] – Pr [X $\leq$ 6], siendo muy cuidadosos al considerar qué resultados han de incluirse y cuáles no (en función de si están o no incluidos los extremos del intervalo original). De este modo, la probabilidad buscada puede calcularse como la diferencia de dos probabilidades obtenidas con *DISTR.BINOM*. En la Ilustración 6 se calcula una de ellas y, en la Ilustración 7, la diferencia entre ambas (con un resultado final de 0,17).

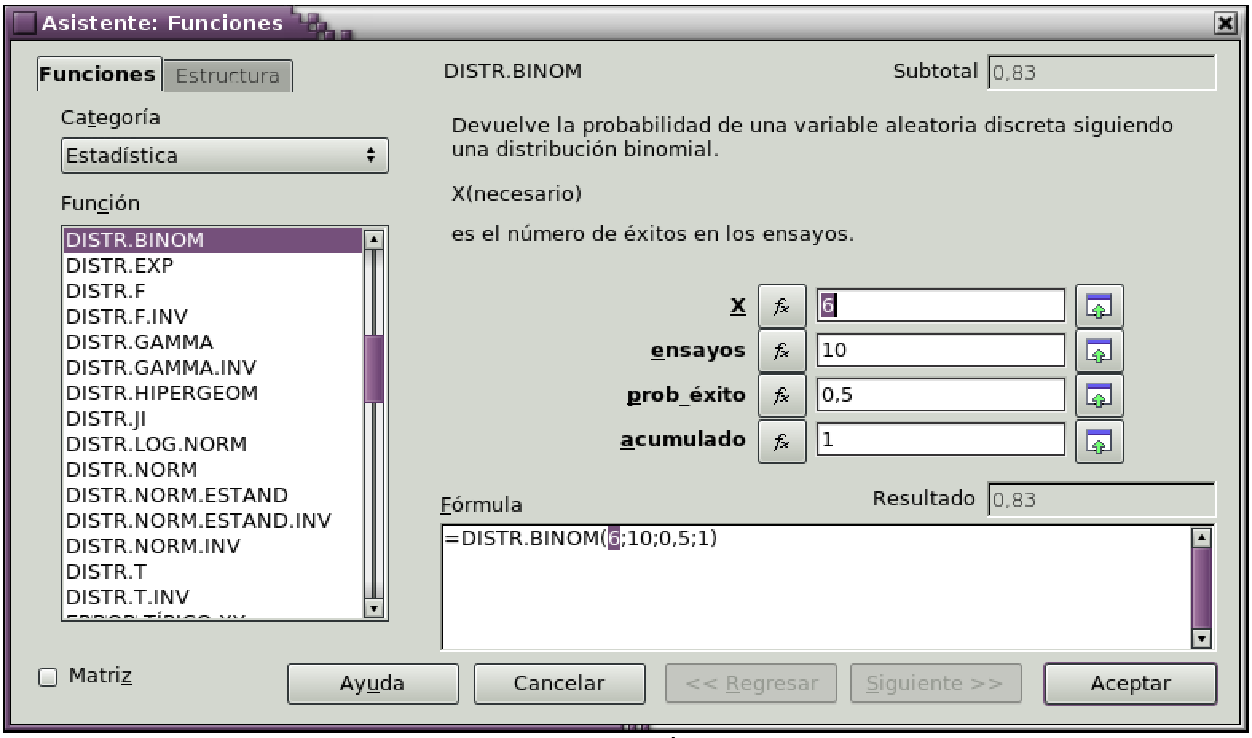

Ilustración 6

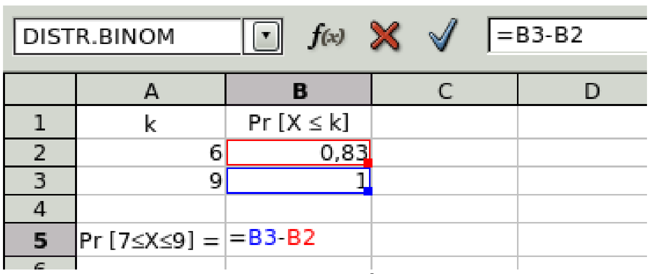

Ilustración 7

De manera análoga procederíamos con la función correspondiente a la **distribución de Poisson**, con la salvedad del parámetro (que ahora es λ, que coincide con la esperanza y con la varianza de la variable.

Ejemplo: *El número medio de automóviles que llega a una gasolinera es de 210* *por hora. Determina la probabilidad de que en un minuto dado lleguen a la gasolinera a lo sumo 10 automóviles y la probabilidad de que el número de automóviles en diez segundos sea mayor que 1 y menor que 7.* Asumiendo que el número de automóviles que llega sigue una distribución de Poisson, al definir X=‘número de automóviles que llega a la gasolinera al minuto’ es claro que X sigue una distribución de Poisson de parámetro λ=210/60 = 3,5. La probabilidad que se pide es Pr [X $\leq$ 10]. En este último caso, tenemos que utilizar el valor *Acumulado* de la función *POISSON* de Calc. El resultado se ofrece en la Ilustración 8.

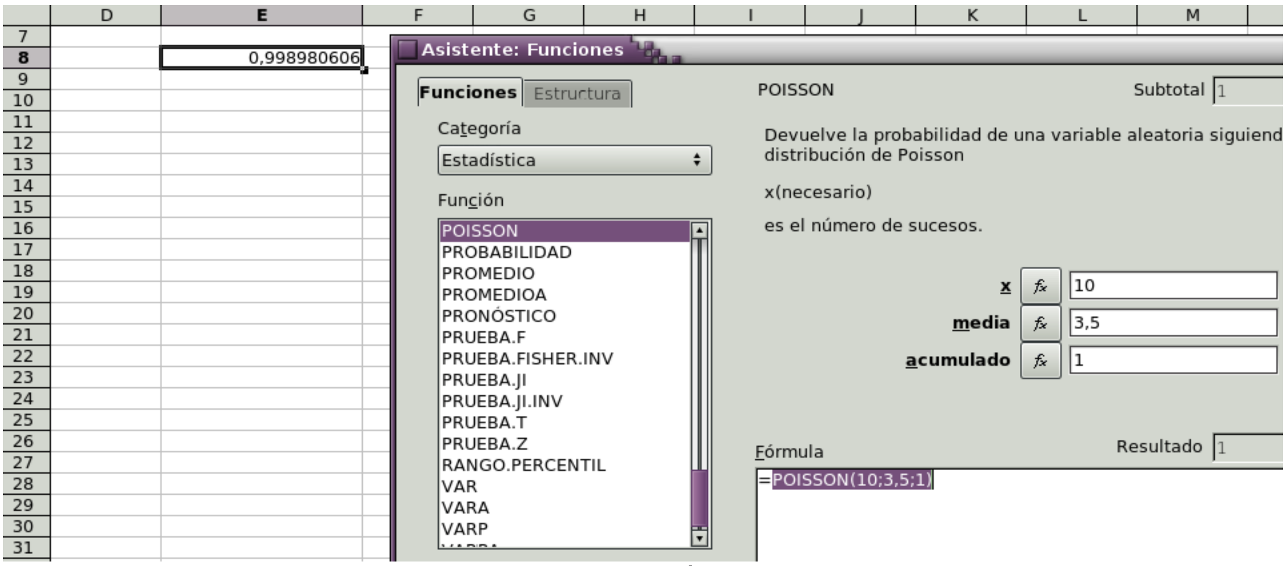

Ilustración 8

Respecto a la segunda probabilidad que se nos pide, si denotamos por Y=‘el número de automóviles que llegan a la gasolinera en diez segundos’, Y sigue una distribución de Poisson de parámetro 3,5 / 6. La probabilidad que se nos pide es Pr [1 < Y < 7]. Para poder calcularla utilizando *POISSON* es necesario escribirla como Pr [1 < Y < 7] = Pr [Y < 7] – Pr [Y $\leq$ 1] = Pr [Y $\leq$ 6] – Pr [Y $\leq$ 1]. En la Ilustración 9 se muestra cómo calcular una de estas dos probabilidades y en la Ilustración 10 la diferencia entre ambas, que ofrece una valor de 0,11644. Como se observa en la Ilustración 9, no es necesario calcular previamente el valor 3,5 / 6 sino que se puede expresar así en la celda correspondiente en la función *POISSON*.

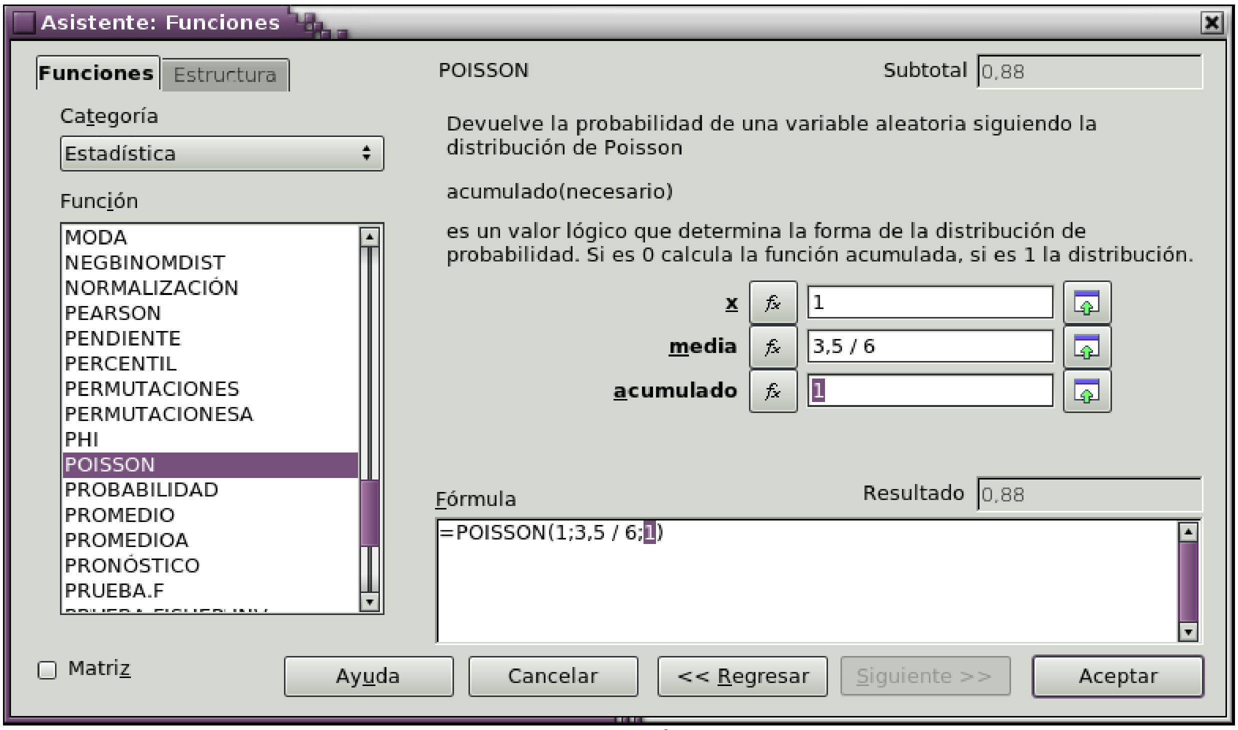

Ilustración 9

	A	B	C
1	k	Pr [X ≤ k]	
2	1	0,88355565	
3	6	0,99999726	
4			
5	Pr [1<X<7] =	=B3-B2	
6			

Ilustración 10

Sea X una variable aleatoria con distribución B (25; 0,32).

Sea Y una variable aleatoria con distribución P (17).

Halla:

a) Pr [X ≤ 10]

b) Pr [X < 14]

c) Pr [X > 14]

d) Pr [X ≥ 10]

e) Pr [5 ≤ Y ≤ 11]

f) Pr [5 ≤ Y < 11]

g) Pr [5 < Y ≤ 11]

h) Pr [5 < Y < 11]

3. Distribuciones continuas

DISTR.NORM (x;μ;σ;1)	Pr [X ≤ x], X → N(μ;σ)
DISTR.NORM.ESTAND (x)	Pr [X ≤ x], X → N(0;1)
DISTR.NORM.INV (p;μ;σ)	x tal que Pr [X ≤ x]=p, X → N(μ;σ)
DISTR.NORM.ESTAND.INV (p)	x tal que Pr [X ≤ x]=p, X → N(0;1)
DISTR.GAMMA (x;α;β;1)	Pr [X ≤ x], X → γ(α;1/β)
DISTR.EXP (x;λ;1)	Pr [X ≤ x], X → Exp(λ)
DIST.WEIBULL (x;α;β;1)	Pr [X ≤ x], X → W(α;β)

Calc dispone también de varias funciones relacionadas con las variables aleatorias continuas, aunque presentan ciertas diferencias con las discretas. La principal diferencia radica en el significado de la casilla de *Acumulado*. En el caso de variables continuas, al poner 'FALSO' en *Acumulado* Calc devuelve el valor de la función de densidad en el punto considerado (que, como es sabido, no se corresponde con la probabilidad de ese punto; dicha probabilidad es siempre cero, por tratarse de una variable continua). Si se pone 'VERDADERO' en *Acumulado*, Calc devuelve (en general, aunque hay algunas excepciones que no veremos aquí) el valor de lo que se denomina función de distribución en el punto k considerado, es decir, Pr [X $\leq k$]. Por lo tanto, dado que la función de densidad como tal sólo la hemos utilizado para calcular probabilidades mediante la integración, en la casilla *Acumulado* pondremos siempre 'VERDADERO'.

Aclarado este punto, el resto del planteamiento es similar al de variables discretas, pues tendremos siempre que expresar la probabilidad que se nos pide en cada problema a partir de probabilidades del tipo Pr [X ≤ k].

Comenzaremos estudiando las distintas funciones relativas a la **distribución normal**. En primer lugar, hay dos funciones *DISTR.NORM* y *DISTR.NORM.ESTAND* que calculan la probabilidad relativa, respectivamente, a una distribución normal cualquiera (indicando la media μ y la desviación típica σ) y a una distribución normal típica (llamada también estándar o cero-uno). Veámoslo con un par de ejemplos.

Ejemplo: *Dada X una variable con distribución N(0;1), calcula la probabilidad de que X sea mayor o igual que 2.* Puesto que hay que calcular Pr [X ≥ 2] y Calc nos da las probabilidades de forma acumulada, pasamos al suceso contrario y obtenemos

Pr [X $\geq$ 2] = 1 – Pr [X < 2]; como estamos trabajando con variables continuas, los "=" se pueden poner o quitar indistintamente (pues la probabilidad de que X sea igual a cualquier número es siempre cero), así que Pr [X $\geq$ 2) = 1 – Pr [X $\leq$ 2], y esta probabilidad ya se puede calcular con Calc mediante la función *DISTR.NORM.ESTAND*. En ella, solamente hay que indicar el valor hasta el cual queremos calcular la probabilidad acumulada, obteniéndose el resultado en la Ilustración 11.

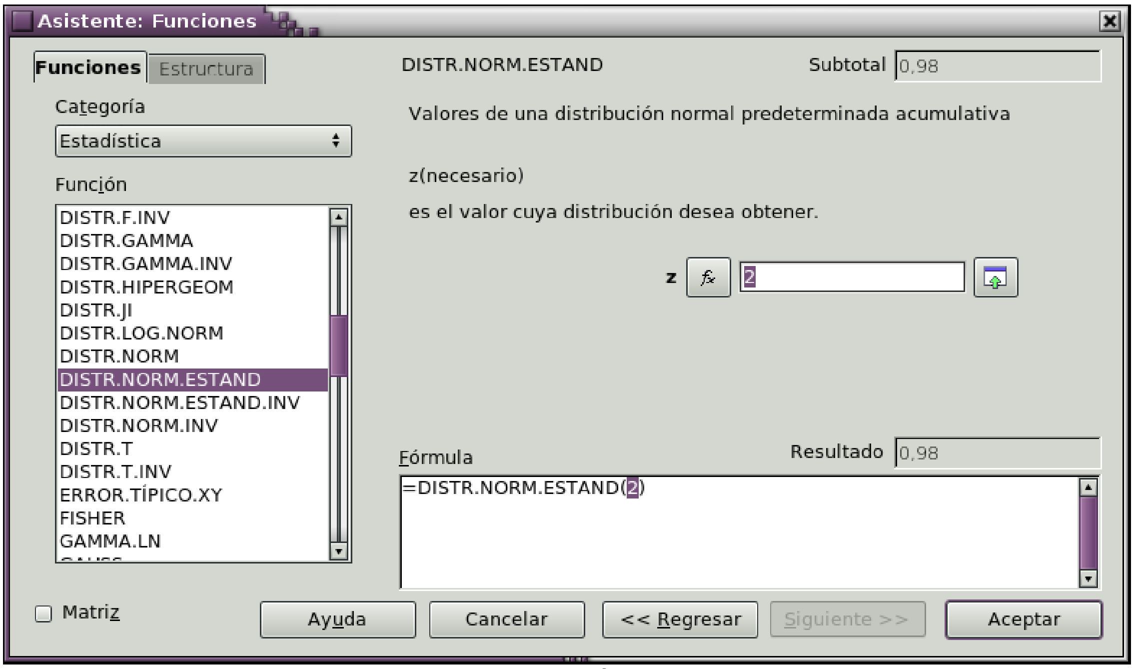

Ilustración 11

Ejemplo: *La ganancia anual, en miles de euros, de una pequeña empresa sigue una distribución normal de media 20 y desviación típica 30. Calcula la probabilidad de que en un año concreto se obtengan más de 30.000 euros de ganancia.* Si X='ganancia anual en miles de euros', se nos pide Pr [X > 30], siendo X una variable con distribución N (20;30). Para adaptar la probabilidad a lo que nos calcula Calc, consideraremos Pr [X > 30] = 1 – Pr [X $\leq$ 30]. Sin necesidad de tipificar, utilizamos la función *DISTR.NORM*, con el valor 30 como parámetro *x*, los valores 20 y 30 como parámetros media y desviación típica, y el valor 1 ó 'VERDADERO' en *Acumulado* para que nos devuelva Pr [X $\leq$ 30]. Restando esta probabilidad a uno, obtenemos el resultado 0,36944 (como se aprecia en la Ilustración 12).

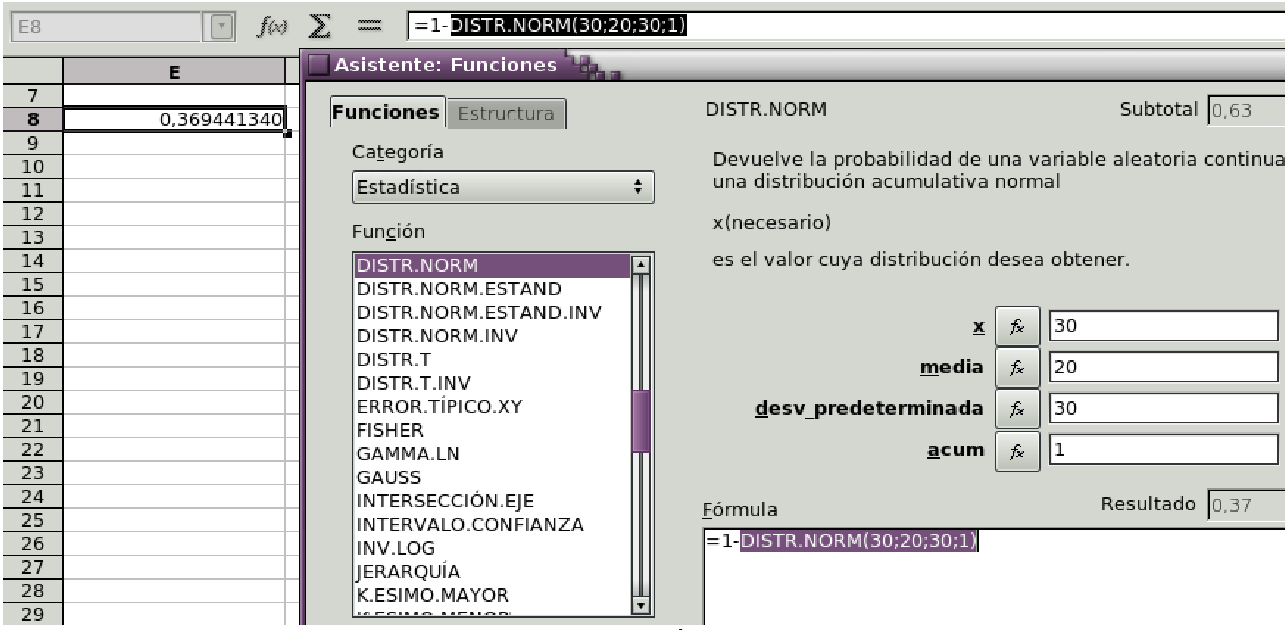

Ilustración 12

Otro par de funciones relacionadas con la distribución normal nos sirven para, dada una probabilidad *p*, calcular el valor k que cumple Pr [X $\leq$ k] = *p*, siendo X una variable normal. Es decir, realizan el cálculo inverso al que realizamos antes (donde, dado un valor, calculábamos su probabilidad). Estas funciones son *DISTR.NORM.INV* (para una normal de parámetros cualesquiera) y *DISTR.NORM.ESTAND.INV* (para la N(0;1)). Su funcionamiento únicamente difiere en que cuando se utiliza la normal típica no es necesario precisar la media ni la desviación típica. Por ello, veremos un ejemplo sólo para la función *DISTR.NORM.INV.*

Ejemplo: *La ganancia anual, en miles de euros, de una pequeña empresa sigue una distribución normal de media 20 y desviación típica 30. Se desea determinar el nivel de ganacias que sólo se superará en el 10% de los años.* Si X=‘ganancia anual en miles de euros’, X sigue una N (20;30) y se nos pide determinar el valor k tal que Pr [X > k] = 0,10, o, equivalentemente, Pr [X $\leq$ k] = 0,9 y así ya tenemos la probabilidad escrita en la forma que nos calcula Calc. Utilizando la función *DISTR.NORM.INV*, obtenemos (ilustración 13) que el valor de k buscado corresponde al número 58,4465.

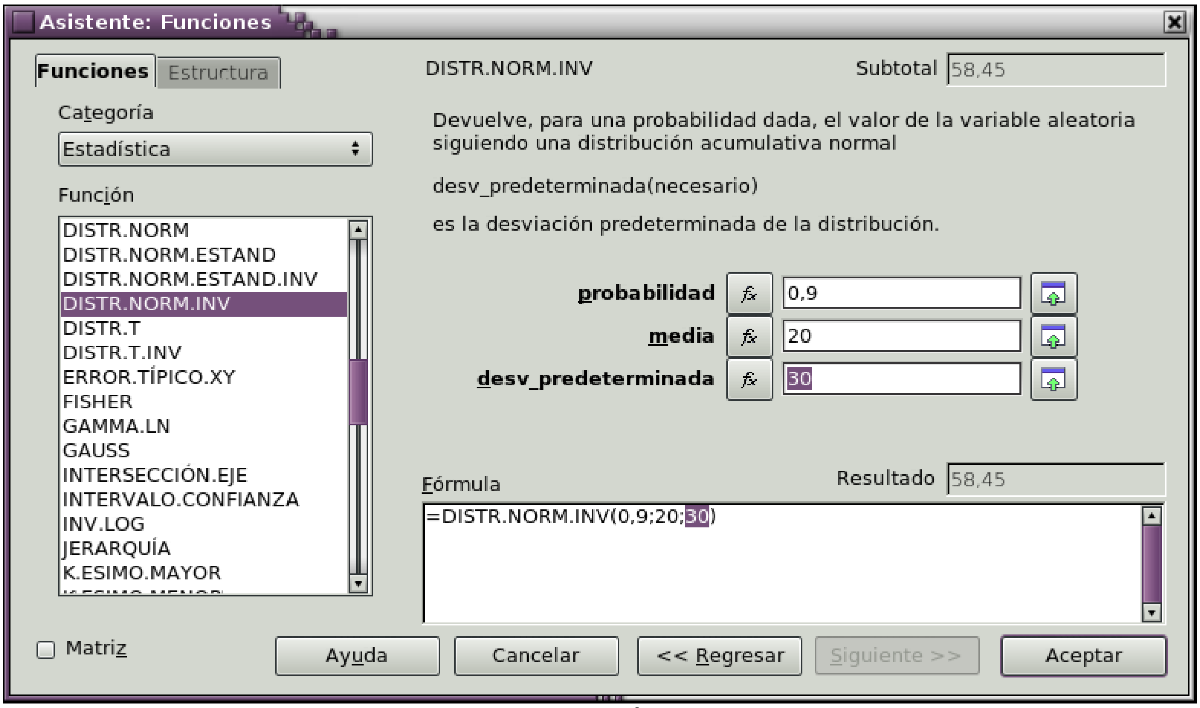

Ilustración 13

Otra de las distribuciones continuas que se han estudiado en la teoría es la **distribución gamma**, y su caso particular que es la **distribución exponencial**. Para ambas existen funciones de Calc que permiten calcular probabilidades relacionadas; son las denominadas *DISTR.GAMMA* y *DISTR.EXP*. Su funcionamiento es análogo a las anteriores, una vez especificados los parámetros de la distribución.

Hay que hacer una advertencia en el caso de la distribución gamma, γ (p;a), puesto que la notación que se ha utilizado en las clases de teoría no coincide exactamente con la que utiliza Calc. Los parámetros que pide Calc para la función *DISTR.GAMMA* son denominados *Alfa* y *Beta*; el primero de ellos, *Alfa*, coincide con el parámetro denotado p en los apuntes teóricos, mientras que el segundo parámetro de Calc, *Beta*, es exactamente 1/a con la notación de los apuntes, por lo tanto, hay que tener cuidado al introducir estos valores. Veámoslo con un ejemplo:

Ejemplo: *La duración de unos determinados fusibles, medida en días, sigue una distribución* γ(100,8). *Calcula la probabilidad de que la duración de un fusible elegido al azar esté entre 11 y 13 días.* Si definimos X=‘duración en días de un fusible’, X sigue una distribución γ(100;8). Se nos pide Pr [11<X<13], que no podemos calcular directamente con Calc, y necesitamos escribirlo como diferencia de probabilidades. Como Pr [$11 < X < 13$] = Pr [$X \leq 13$] – Pr [$X \leq 11$] podemos obtener la probabilidad que nos piden como la diferencia entre dos probabilidades que sí calcularemos con Calc. En la Ilustración 14 se recoge solamente el cálculo de la probabilidad del 11 pues la correspondiente a 13 sería similar. Como hemos comentado, hay que tener cuidado al teclear el segundo parámetro, en la casilla *Beta*, pues para Calc no es 8, sino su inverso, 1/8 (es suficiente con teclearlo así, Calc realiza el cálculo automáticamente). En la Ilustración 15 mostramos el cálculo de la diferencia, restando las dos celdas que contienen las probabilidades mediante una fórmula y obteniendo un resultado final de Pr [11 < X < 13] = 0,5541.

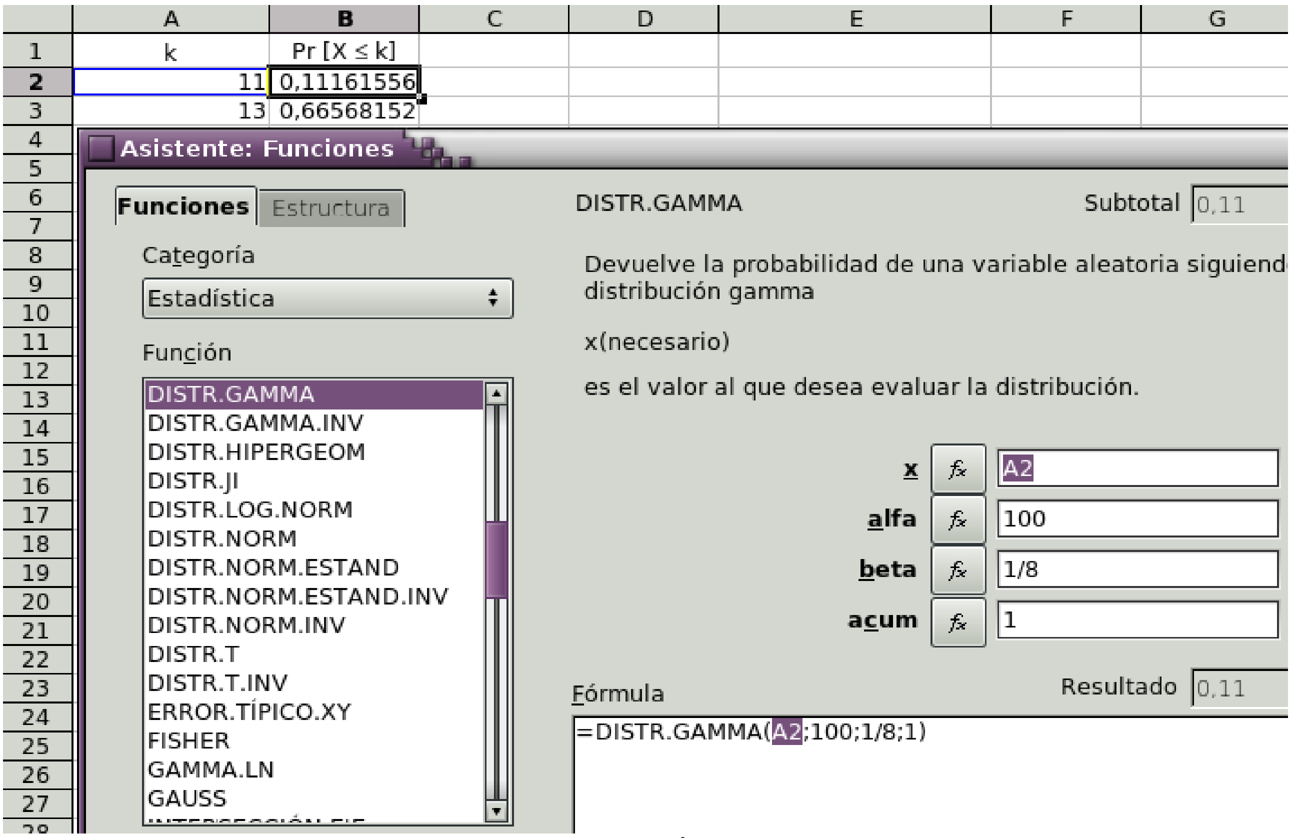

Ilustración 14

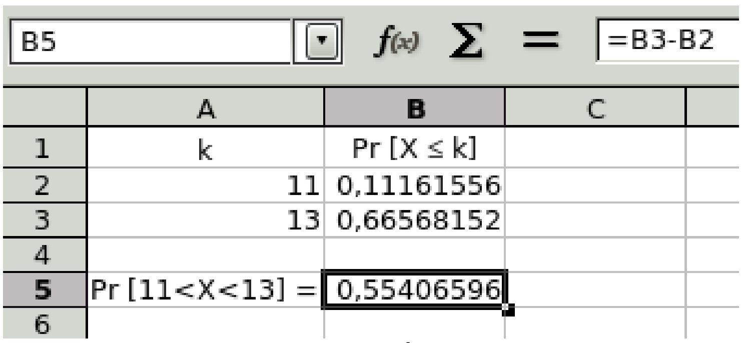

Ilustración 15

Con la distribución exponencial no ocurre esto, pues la notación que utiliza Calc en la función *DISTR.EXP* coincide con la utilizada en la teoría: se denota el parámetro como *Lambda* (nombre de la letra griega λ).

Ejemplo: *Se estudia la duración (en días) de un cierto tipo de pilas y resulta seguir una distribución exponencial de parámetro 0,1. Calcula la probabilidad de que una pila elegida al azar dure menos de 10 días.* Si denotamos X=‘duración en días de una pila’ se tiene que X sigue una distribución *exp*(0,1). La probabilidad que se nos pide es Pr [X < 10] = Pr [X ≤ 10], luego puede ser calculada directamente con Calc mediante la función *DISTR.EXP*. El resultado se aprecia en la ilustración 16.

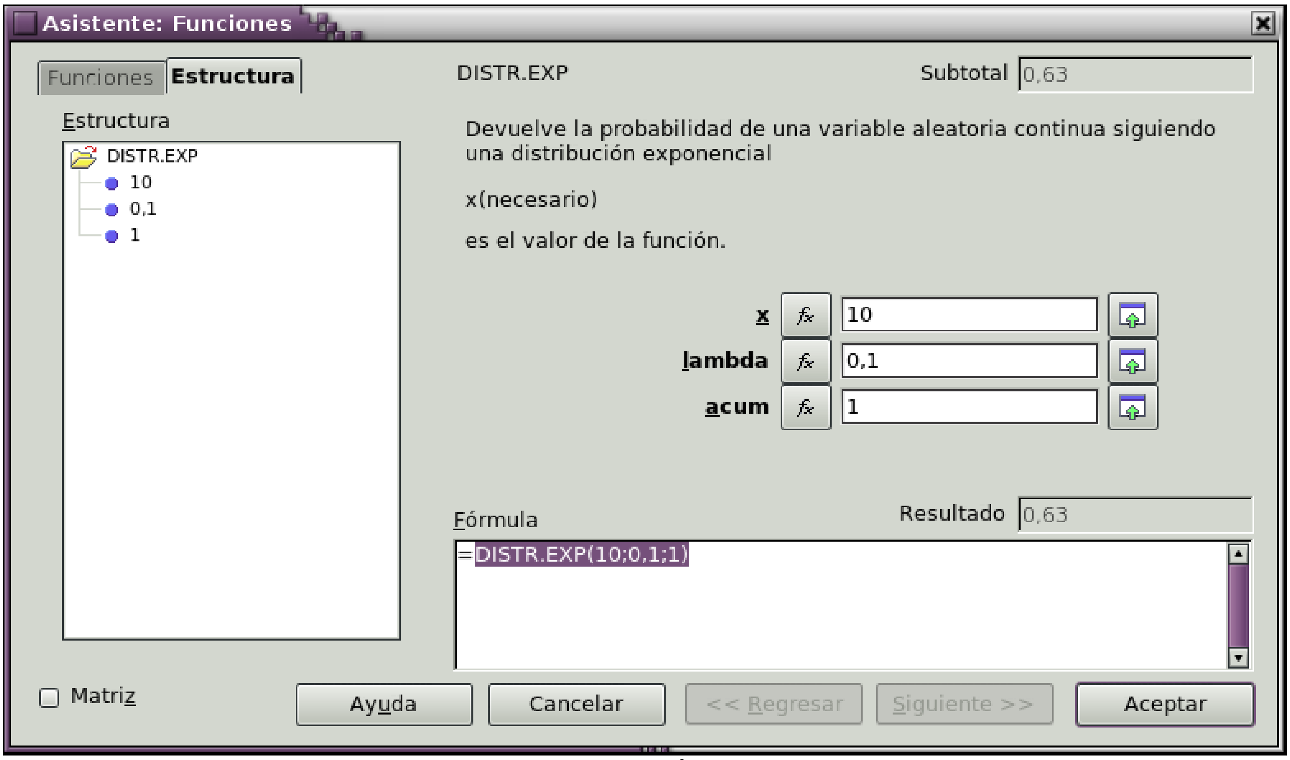

Ilustración 16

Finalmente, mencionaremos también la **distribución de Weibull**, que tiene una función de Calc asociada llamada *DIST.WEIBULL*, y que utiliza casi la misma notación para sus parámetros que la que se ha visto en teoría: el primer parámetro se denota por α (*Alfa*) en ambos sitios, mientras que el segundo parámetro, θ en los apuntes, se denota *Beta* en Calc.

Ejemplo: *La duración (medida en miles de horas) de un determinado tipo de componente mecánica sigue una distribución de Weibull de parámetros θ =* $\sqrt{50}$ *y α = 2. Halla la probabilidad de que una de esas componentes trabaje de forma correcta más de 10000 horas.* Si llamamos X=‘duración en miles de horas de una componente mecánica’, X sigue una distribución Weibull (2, $\sqrt{50}$) y se nos pide Pr [X>10], que calcularemos como $\Pr[X>10] = 1 - \Pr[X \leq 10]$. Por lo tanto, con la función *DIST.WEIBULL* obtendremos el resultado que se aprecia en la Ilustración 17. Como se observa, no es necesario calcular previamente el valor $\sqrt{50}$ y, además, realizamos en la misma celda el cálculo $1 - \Pr[X \leq 10]$ para obtener un resultado final de $\Pr[X>10] = 0{,}1353$.

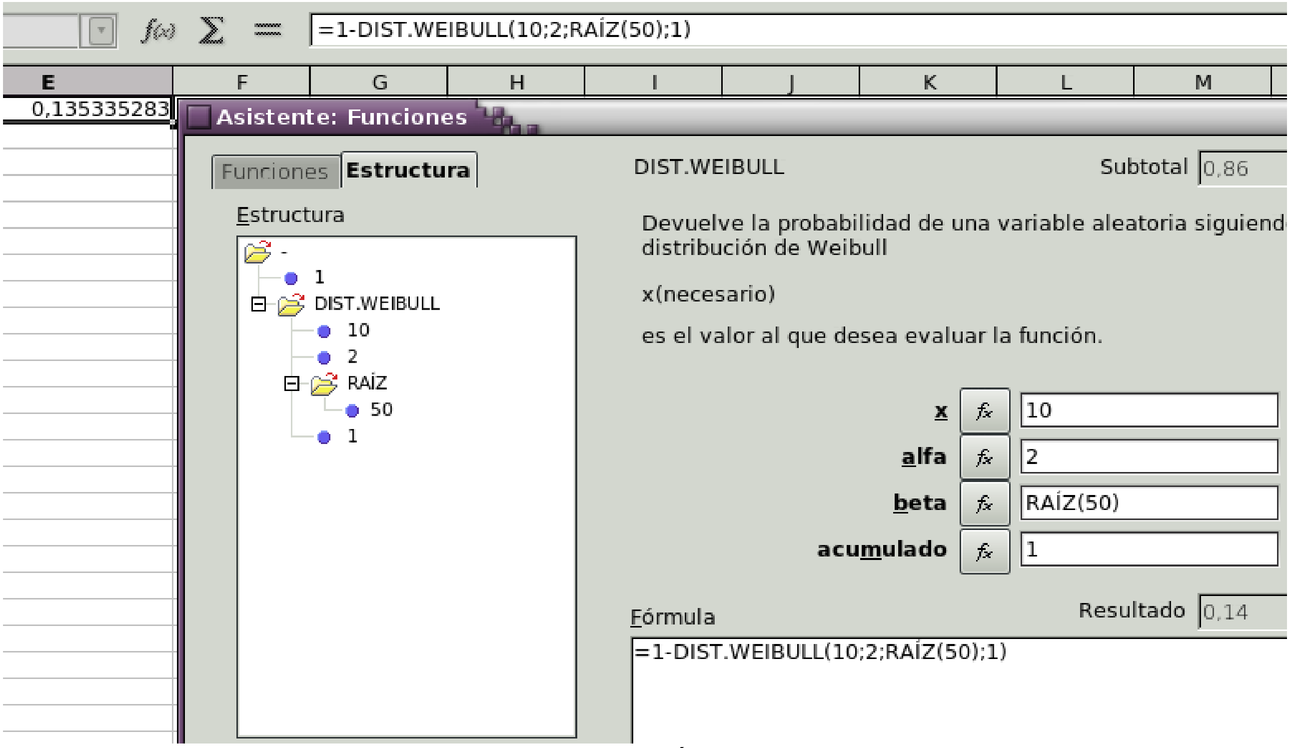

Ilustración 17

Sea U una variable aleatoria con distribución N (25; 5,8).

Sea V una variable aleatoria con distribución Exp (0,05).

Sea X una variable aleatoria con distribución γ (40; 2).

Sea Y una variable aleatoria con distribución W (2,5; 17).

Halla:

i) $\Pr[U \leq 21]$

j) $\Pr[U < 21]$

k) $\Pr[U > 15]$

l) $\Pr[U \geq 15]$

m) $\Pr[14 \leq U \leq 20]$

n) $\Pr[14 \leq U < 20]$

o) $\Pr[14 < U \leq 20]$

p) $\Pr[14 < U < 20]$

q) $\Pr[14 < V < 20]$

r) $\Pr[14 < X < 20]$

s) $\Pr[14 < Y < 20]$

t) k tal que $\Pr[U < k] = 0{,}2$

u) k tal que $\Pr[U > k] = 0{,}35$

v) k tal que $\Pr[|U-25| < k] = 0{,}9$

4. EJERCICIOS ADICIONALES

1. La proporción de individuos en una población con renta anual superior a 240 000 euros es 0,005%. Determina la probabilidad de que entre 5000 individuos elegidos al azar haya 2 con ese nivel de renta y la probabilidad de que haya más de 2.
2. Una empresa manufacturera elabora diariamente una cantidad de producto que (en toneladas) sigue una distribución γ(3;1). Calcula la probabilidad de que, en una día cualquiera, la empresa produzca más de cuatro toneladas.
3. En un laboratorio en el cual se analizan muestras de agua mineral, el pH de las muestras sigue una distribución que puede considerarse normal de media 6,5 y desviación típica 0,5. Elegida una muestra al azar, ¿cuál es la probabilidad de que su pH sea exactamente igual 7? ¿Y la probabilidad de que supere 7?
4. La calificación media de un examen ha sido de 5,5, con una desviación típica de 1,5 y el conjunto de notas se ajusta a una distribución normal. El profesor quiere calificar con sobresaliente al 10% de la clase. ¿A partir de qué nota se conseguirá el sobresaliente?
5. Si el tiempo (en meses) que un vial de suero salino se mantiene sin alteraciones tras su envasado sigue una distribución exponencial de parámetro 0,02, ¿cuál es la probabilidad de que un vial de suero dure más de 5 meses sin alteraciones?

Soluciones

a) 0,857626	f) 0,048940	k) 0,957659	p) 0,165381	u) 27,234859
b) 0,988818	g) 0,083995	l) 0,957659	q) 0,128706	v) 9,540151
c) 0,003618	h) 0,048450	m) 0,165381	r) 0,502041	
d) 0,255488	i) 0,245206	n) 0,165381	s) 0,317542	
e) 0,084484	j) 0,245206	o) 0,165381	t) 20,118597	

1. Denominando X al número de individuos en la muestra de 5 000 con renta superior a 240 000 euros, el problema se puede resolver mediante la distribución exacta de X, que sería una <u>binomial</u> con n = 5 000 y p = 0,000 05 (función *DISTR.BINOM*). También se puede utilizar la distribución aproximada (ley de sucesos raros), ya que $n > 50$, $p < 0{,}01$ y $np = 0{,}25 < 5$ y, por lo tanto, X sigue aproximadamente una distribución de <u>Poisson</u> de parámetro np = 0,25 (función *POISSON*). En cualquier caso, se pide Pr [X = 2] = 0,024 334 9 (0,024 337 5 con la aproximación) y Pr [X > 2] = 1 – Pr [X $\leq$ 2] = 1 – 0,997 84 = 0,002 16 (con la aproximación se obtiene 0,002 161).
2. Si X es la cantidad de toneladas producidas diariamente, que sigue una distribución gamma de parámetros 3 y 1, se nos pide Pr [X > 4] = 1 – Pr [X $\leq$ 4]. En este caso el segundo parámetro es un 1, por lo que no es necesario calcular su inverso para

utilizar la función *DISTR.GAMMA*, y obtenemos un resultado de Pr [X > 4] = 1 – 0,7619 = 0,2381.

3. Evidentemente, si denotamos por X el pH de las muestras, X sigue una distribución N(6,5;0,5) y, por consiguiente, la probabilidad de que sea exactamente igual a 7 es 0, por ser una variable continua. No ocurre así con la otra probabilidad pedida, que es Pr [X > 7]; como hemos hecho ya en varias ocasiones, la calcularemos como Pr [X > 7] = 1 – Pr [X ≤ 7]; utilizando la función *DISTR.NORM* obtenemos Pr [X > 7] = 0,1586.
4. Denotando por X la calificación del examen, su distribución es N(5,5; 1,5), si el profesor quiere poner sobresaliente al 10% de la clase, la nota de corte para el sobresaliente, S, debe ser tal que se obtenga una nota superior o igual con probabilidad 0,1, es decir, Pr [X ≥ S] = 0,1. Para escribir esta probabilidad en forma de menor o igual, pasamos al suceso contrario y obtenemos Pr [X ≤ S] = 0,9, de modo que, utilizando la función *DISTR.NORM.INV* obtendremos el valor que deja probabilidad 0,9 a la izquierda en la normal de los parámetros especificados, que es 7,42.
5. Llamando X a la duración sin alteraciones en meses de un vial de suero salino, esta variable sigue una distribución *exp*(0,02). Se pide calcular Pr [X > 5] = 1 – Pr [X ≤ 5]. Mediante la función *DISTR.EXP* obtenemos el resultado 0,9048.