Práctica 5

Modelos de distribuciones

5.1. Introducción

En los menús `Distribuciones`→`Distibuciones continuas/Distribuciones discretas` encontramos un conjunto de opciones que permiten representar las funciones de densidad y de probabilidad de diversos modelos de distribuciones continuas y discretas, así como la correspondiente función de distribución. También es posible calcular probabilidades, cuantiles y generar valores aleatorios de acuerdo a estos modelos.

Para los **modelos discretos** que aparecen en R-Commander disponemos de las siguientes opciones:

- `Cuantiles (modelo)`: dada una variable aleatoria X discreta, el cuantil de orden k (opción cola izquierda) es el menor valor c_k de la variable para el cual se verifica que $P(X \leq c_k) \geq k$.
- `Probabilidades (modelo) acumuladas`: función de distribución del modelo considerado. Para cada valor x la opción `cola izquierda` muestra $P[X \leq x]$. La opción `cola derecha` muestra $P[X > x]$.
- `Probabilidades (modelo)`: muestra la probabilidad de los valores que toma la variable.
- `Gráfica de la distribución (modelo)`: representa la función de probabilidad o de distribución del modelo considerado.
- `Muestra de una distribución (modelo)`: genera valores aleatorios de acuerdo al modelo considerado.

Para las **distribuciones continuas** disponemos de las siguientes opciones:

- `Cuantiles (modelo)`: el cuantil k, $0 \leq k \leq 1$, de una variable aleatoria X continua es el valor c_k para el cual $P(X \leq c_k) = k$. Nótese que $P(X \geq c_k) = 1-k$.

- `Probabilidades (modelo) acumuladas`: función de distribución del modelo considerado. La opción `cola izquierda` del menú proporciona, para cada valor x, el valor de $P[X \leq x] = P[X < x]$. La opción `cola derecha` proporciona $P[X > x] = P[X \geq x] = 1 - P[X \leq x]$.

- `Gráfica de la distribución (modelo)`: representa gráficamente la función de densidad o de distribución del modelo considerado.

- `Muestra de una distribución (modelo)`: genera valores aleatorios de acuerdo al modelo considerado.

5.2. Modelo binomial

Un examen tipo test consta de 21 preguntas, cada una de las cuáles tiene 4 posibles respuestas. Si un estudiante responde a todas las preguntas al azar,

a) Determinar la distribución del número de preguntas respondidas correctamente por el estudiante.

   Cada pregunta tiene cuatro posibles respuestas y sólo una de las respuestas es correcta por lo que, si una pregunta se responde al azar, la probabilidad de acertarla es 0.25. Además el examen está formado por 21 preguntas y cada una se acierta o no con independencia de las demás por lo que el número de respuestas correctas será una variable aleatoria $X \sim B(21, 0.25)$.

b) Representar las funciones de probabilidad y de distribución de esta variable.

   Hacemos las representaciones mediante el menú

   `Distribuciones`→`Distribuciones discretas` →`Distribución binomial`→
   →`Gráfica de la distribución binomial`

   En la ventana emergente indicamos: `Ensayos binomiales: 21` y `Probabilidad de éxito:` 0.25.

   Nótese que, al representar la función de distribución, aparecen líneas verticales en los `puntos de salto` que realmente no forman parte de la representación de la función.

c) Calcular la probabilidad de que el estudiante responda de manera correcta a, exactamente, 6 preguntas.

   Usamos el menú: `Distribuciones`→`Distribuciones discretas` →`Distribución binomial`→ `Probabilidades binomiales`.

   Introducimos los parámetros de la distribución en la ventana emergente y pulsamos en el botón `Aceptar`. Observamos que el programa devuelve las probabilidades de que la variable tome cada uno de sus posibles valores. De este modo obtenemos que $P(X = 6) = 0.1770398$.

d) Calcular la probabilidad de aprobar el examen si para ello es necesario responder correctamente a, al menos, 11 preguntas.

   Tenemos que calcular $P(X \geq 11)$.

   Usamos el menú: `Distribuciones`→`Distribuciones discretas` →`Distribución binomial`→ `Probabilidades binomiales acumuladas`.

   En la ventana emergente observamos dos opciones cola izquierda y cola derecha. Podemos calcular la probabilidad solicitada usando la opción `cola derecha`. En ese caso, puesto que $P(X \geq 11) = P(X > 10)$ indicamos como `Valor de la variable` el valor 10, `Ensayos binomiales` 21 y `Probabilidad de éxito` 0.25. Obtenemos como solución el valor 0.00642271.

   También podemos calcular la probabilidad solicitada mediante la opción `cola izquierda`. En ese caso, necesitamos escribir la probabilidad en función del suceso complementario. De este modo, $P(X \geq 11) = 1 - P(X < 11) = 1 - P(X \leq 10)$. En la ventana emergente indicamos como `Valor de la variable` el valor 10 y obtenemos que $P(X \leq 10) = 0.9935773$. Por lo tanto $P(X \geq 11) = 1 - P(X \leq 10) = 1 - 0.9935773 = 0.0064227$.

   Recuérdese que si la variable X es discreta, indicando como `valor de la variable` el valor k, la opción `cola izquierda` calcula $P(X \leq k)$ y la opción `cola derecha` calcula $P(X > k)$.

5.3. Distribución de Poisson

El número de accesos por minuto a una página web sigue una distribución de Poisson de media 5.

a) Representar las funciones de probabilidad y de distribución de esta variable.

   Para representar ambas funciones utilizamos el menú `Distribuciones`→`Distribuciones discretas` →`Distribución de Poisson`→`Gráfica de la distribución de Poisson`.

b) Determinar la probabilidad de que, en un minuto, se produzcan exactamente 4 accesos.

   El número de accesos que se producen en un minuto es una variable aleatoria $X \sim \mathcal{P}(5)$.

   Usamos el menú: `Distribuciones`→`Distribuciones discretas` →`Distribución de Poisson`→ `Probabilidades de Poisson`. En la ventana emergente introducimos como `media de la variable` el valor 5 y pulsamos el botón `Aceptar`.

   Podemos observar que R nos devuelve $P(X = k), k = 0, \ldots, 14$. Aunque X puede tomar cualquier valor entero positivo, no se muestran las probabilidades de los valores mayores o iguales que 15 por ser muy pequeñas. Obtenemos entonces que $P(X = 4) = 0.1754673698$.

c) Determinar la probabilidad de que, en dos minutos, el número de accesos sea, a lo más, de 7 (supondremos que cada acceso se produce de manera independiente de los demás).

Consideramos las variables X_1: número de accesos en el minuto 1 y X_2: número de accesos en el minuto 2. Entonces ambas variables son independientes y, puesto que $X_i \sim \mathcal{P}(5), i = 1, 2$, el número total de accesos es $X_1{+}X_2 \sim \mathcal{P}(5{+}5) = \mathcal{P}(10)$.

Calculamos entonces $P(X_1{+}X_2 \leq 7)$. Para ello usamos el menú `Distribuciones`→ →`Distribuciones discretas` →`Distribución de Poisson`→ `Probabilidades de Poisson acumuladas`. En la ventana emergente introducimos como `valor de la variable` 7, como `Media` el valor 10 y seleccionamos la opción `cola izquierda`. Obtenemos entonces que $P(X_1 + X_2 \leq 7) = 0.2202206$.

d) Nos preguntan por el número de accesos que, como máximo, se producirán en el próximo minuto. Determinar el menor valor que debemos dar como respuesta si deseamos acertar con probabilidad no inferior a 0.9.

Consideramos nuevamente la variable X: número de accesos que se producen en un minuto. Nos preguntan por el menor valor, a, del número de accesos que verifica que $X \leq a$ con probabilidad mayor o gual que 0.9. Esto es, el menor a para el que se verifica $P(X \leq a) \geq 0.9$. Nótese que, de acuerdo a la definición que se dio en la sección 5.1, a es el cuantil 0.9 de la distribución de X (*opción cola izquierda*).

Para calcularlo usamos el menú `Distribuciones`→`Distribuciones discretas` →`Distribución de Poisson`→ `Cuantiles de Poisson`. En la ventana emergente introducimos en el campo `Probabilidades` el valor 0.9 y en el campo `Media` el valor 5 (recuérdese que $X \sim \mathcal{P}(5)$). Finalmente elegimos la opción `Cola izquierda` y pulsamos el boton `Aceptar`.

Obtenemos entonces que, con probabilidad no inferior 0.9, podemos afirmar que el número de accesos que como máximo se producirán en el próximo minuto es 8.

5.4. Distribuciones geométrica y binomial negativa

Representar la función de probabilidad de una distribución geométrica de parámetro 0.25. Deducir la relación que existe entre la definición, vista en teoría, de la distribución geométrica y la definición que utiliza `R`.

Para representar la función de probabilidad indicada usamos el menú `Distribuciones`→`Distribuciones discretas` →`Distribución geométrica`→ `Gráfica de la distribución geométrica`. Podemos observar que distribución representada por R toma valores enteros mayores o iguales que 0 mientras que la distribución geométrica definida en la teoría toma valores mayores o iguales que 1. Esto es debido a que la distribución geométrica se ha definido en la teoría como el número de experimentos de Bernouilli independientes que se hacen hasta obtener el primer éxito mientras que R la define como el número de experimentos que se hacen antes del primer éxito.

Consideremos ahora un modelo de examen formado por preguntas tipo test, con cuatro posibles respuestas, y supongamos que al alumno se le hace un número indefinido de preguntas que va respondiendo al azar. Responder a las siguientes cuestiones:

a) Determinar la probabilidad de que el alumno dé su primera respuesta correcta, en la cuarta pregunta que responde.

   Consideramos la variable X : número de preguntas que responde el alumno hasta dar la primera respuesta correcta. Entonces $X \sim Ge(0.25)$ y tenemos que calcular $P(X = 4)$. Para ello usamos el menú `Distribuciones`→`Distribuciones discretas` →`Distribución Geométrica`→ `Probabilidades Geométricas`.

   En la ventana emergente indicamos la probabilidad éxito del experimento de Bernouilli asociado, que es 0.25, y pulsamos el botón `Aceptar`.

   R define la distribución geométrica como el número de ensayos de una secuencia experimentos de Bernouilli que se realizan antes del primer éxito. Por lo tanto, $P(X = 4)$ corresponde con la probabilidad que se encuentra, en la salida de R-Commander, junto al valor 3. Por lo tanto, $P(X = 4) = 0.1054687500$.

b) Determinar la probabilidad de que necesite responder, como mucho, a 6 preguntas hasta dar una respuesta correcta.

   Tenemos que calcular, en este caso, $P(x \leq 6)$. Usamos el menú `Distribuciones`→ →`Distribuciones discretas` →`Distribución Geométrica`→ `Probabilidades Geométricas acumuladas`.

   Responder como mucho 6 preguntas hasta dar una respuesta correcta es equivalente a responder, como mucho, 5 preguntas antes de dar una respuesta correcta. Por lo tanto, en la ventana emergente, indicamos, en el campo `Valores de la variable`, el valor 5. Como probabilidad de éxito ponemos 0.25 y seleccionamos la opción `Cola izquierda`. Obtenemos entonces que la probabilidad buscada es 0.8220215.

c) Determinar la probabilidad de que el alumno dé su tercera respuesta correcta en la sexta pregunta que responde.

   Consideramos la variable Y : número de preguntas que debe responder el alumno hasta dar la tercera respuesta correcta. Entonces, $Y \sim B^{-}(3, 0.25)$ y tenemos que calcular $P(Y = 6)$.

   Usamos el menú `Distribuciones`→`Distribuciones discretas` →`Distribución binomial negativa`→ `Probabilidades binomiales negativas`.

   Los parámetros que se deben considerar para hacer los calculos correspondientes a la distribución binomial negativa, $B^{-}(r, p)$, son:

   - `Número de éxitos`: número de éxitos que deben alcanzarse en la secuencia de experimentos de Bernouilli. Corresponde con el parámetro r de la distribución.

- `Probabilidad de éxito`: es la probabilidad de éxito en el experimento de Bernouilli asociado al experimento aleatorio. Corresponde con el parámetro p.
- `Valor de la variable`: al solicitar probabilidades de una variable $X \sim B^{-}(r,p)$, hay que tener en cuenta que R define $P(X=k)$, $k \in \{0,1,2,\ldots\}$, como la probabilidad de que se produzcan k fracasos antes del r-ésimo éxito.

Así pues, en la ventana emergente, indicamos `Número de éxitos:` 3 y `Probabilidad de éxito:` 0.25 y pulsamos el botón `Aceptar`.

R define el modelo binomial negativo como el número de fracasos antes del r-ésimo éxito. Por otra parte, si el tercer acierto del alumno se produce en el sexto acierto, antes ha fallado en tres ocasiones. Así pues, para calcular $P(Y=6)$ buscamos en la salida de R la probabilidad asociada al valor 3. De este modo encontramos que $P(Y=6)=0.0659179688$.

5.5. Distribución hipergeométrica

Un lote de 20 dispositivos electrónicos contiene cinco dispositivos defectuosos. Si se eligen 10 dispositivos al azar, ¿cuál es la probabilidad de que 2 de los 10 dispositivos estén defectuosos? (Sol: 0.34829721).

Nota: Si X sigue una distribución hipergeométrica, R define $P(X=t)$ como la probabilidad de encontrar t *bolas blancas* entre todas las bolas elegidas al azar.

Consideramos la variable aleatoria: X : número de dispositivos, de los 10 elegidos, que están defectuosos, para la que se cumple que $X \sim H(20,5,10)$. Seleccionamos el menú `Distribuciones`$\rightarrow$`Distribuciones discretas` $\rightarrow$`Distribución hipergeométrica`$\rightarrow$ `Probabilidades hipergeométricas`. Obtenemos que la probabilidad buscada es 0.34829721.

5.6. Distribuciones exponencial y gamma

El tiempo de funcionamiento sin fallos de un ordenador sigue una variable aleatoria con distribución exponencial de media 20 minutos.

a) Representar las funciones de densidad y de distribución de esta variable.

   Hacemos la representación mediante el menú

   `Distribuciones`$\rightarrow$`Distribuciones continuas` $\rightarrow$`Distribución exponencial`$\rightarrow$
   $\rightarrow$`Gráfica de la distribución exponencial`.

b) Determinar la probabilidad de que el tiempo de funcionamiento sin fallos sea, al menos, 22 minutos.

Consideramos la variable aleatoria T: tiempo de trabajo sin fallos del ordenador. Entonces $T \sim Exp(1/20) = Exp(0.05)$ y tenemos que calcular $P(T \geq 22)$. Usamos el menú:

`Distribuciones`$\rightarrow$`Distribuciones continuas` $\rightarrow$`Distribución exponencial`$\rightarrow$ $\rightarrow$`Probabilidades exponenciales acumuladas`.

En la ventana emergente indicamos `Valor de las variables:` 22 y `Parámetro de la exponencial:` 0.05. Al ser T una variable continua se verifica que $P(T \geq 22) = P(T > 22)$, por lo que marcamos la opción cola derecha y obtenemos que $P(T > 22) = 0.3328711$.

c) Determinar un periodo de tiempo tal que, con probabilidad 0.99, podamos afirmar que el ordenador trabajará sin fallos durante ese periodo.

Si en el intervalo $[0, t]$ no se producen fallos entonces $T > t$. Por lo tanto, buscamos un valor t tal que $P(T > t) = 0.99$ o, lo que es igual, $P(T \leq t)$=0.01. Dicho valor es el cuantil 0.01 de la distribución de T. Para calcularlo, usamos el menú

`Distribuciones`$\rightarrow$`Distribuciones continuas` $\rightarrow$`Distribución exponencial`$\rightarrow$ $\rightarrow$`Cuantiles exponenciales`. En la ventana emergente indicamos, `Probabilidades:` 0.01, `Parámetro de la exponencial:` 0.05 y seleccionamos la opción `Cola izquierda`. Obtenemos entonces que $t = 0.2010067$ y el intervalo buscado, en minutos, es $[0, 0.2010067]$.

d) Determinar la probabilidad de que el tiempo entre cuatro fallos consecutivos sea inferior a 60 minutos. (Sol: 0.5768099).

Nota: si $X \sim G(a, p)$, entonces el parámetro de forma al que hace referencia R es p y el parámetro de escala es $1/a$.

Sabemos que el tiempo de trabajo sin fallos o, equivalentemente, el tiempo que transcurre entre dos fallos consecutivos sigue una distribución $Exp(0.05)$. Entonces el tiempo que transcurre entre cuatro fallos consecutivos viene dado por $T_1 + T_2 + T_3$, donde cada $T_i, i = 1, 2, 3$ es el tiempo que transcurre entre el i-ésimo fallo y el siguiente. Además, T_1, T_2 y T_3 son variables independientes que siguen una distribución $Exp(0.05)$.

Por ser T_1, T_2 y T_3 variables independientes, se verifica que $T_1 + T_2 + T_3 \sim G(0.05, 3)$. Tenemos que calcular entonces $P(T_1 + T_2 + T_3 \leq 60)$. Para ello usamos el menú `Distribuciones`$\rightarrow$`Distribuciones continuas` $\rightarrow$`Distribución gamma`$\rightarrow$`Probabilidades acumuladas`. En la ventana emergente indicamos `Valor de la variable:` 60, `Parámetro de forma`: 3, `Parámetro de escala:` 20. Obtenemos que la probabilidad buscada es 0.5768099.

5.7. Distribución normal

El contenido de zumo en litros, que un proceso de llenado automático deposita en las botellas, sigue una distribución normal de media 2 y desviación típica 0.1.

a) Representar las funciones de densidad y de distribución de esta variable.

`Distribuciones`→`Distribuciones continuas` →`Distribución normal`→ →`Gráfica de la distribución normal`.

En lo que sigue trabajaremos con la variable aleatoria X: cantidad de zumo, en litros, que contiene una botella.

b) Determinar la probabilidad de que una botella elegida al azar contenga más de 1.9 litros de zumo.

Tenemos que calcular $P(X > 1.9)$. Seleccionamos el menú `Distribuciones`→`Distribuciones continuas`→`Distribución normal`→`Probabilidades normales acumuladas`.

En la ventana emergente indicamos como `Valor de la variable` 1.9, `Media` 2, `Desviación típica` 0.1 y seleccionamos la opción `Cola derecha`. Obtenemos que $P(X > 1.9) = 0.8413447$.

c) Determinar la probabilidad de que una botella elegida al azar contenga entre 1.95 y 2.1 litros de zumo.

Calculamos ahora $P(1.95 \leq X \leq 2.1)$. Por ser X una variable continua,

$P(1.95 \leq X \leq 2.1) = F_X(2.1) - F_X(1.95) = P(X \leq 2.1) - P(X \leq 1.95)$.

Usamos el menú `Distribuciones`→`Distribuciones continuas` →`Distribución normal`→`Probabilidades normales acumuladas`.

Para calcular $P(X \leq 2.1)$ introducimos los siguientes valores `Valores de la variable` 2.1, `Media` 2, `Desviación típica` 0.1 y marcamos la opción `Cola izquierda`. Obtenemos entonces que $P(X \leq 2.1) = 0.8413447$. Repetimos el procedimiento para calcular que $P(X \leq 1.95) = 0.3085375$.

Por lo tanto, $P(1.95 \leq X \leq 2.1) = 0.8413447 - 0.3085375 = 0.5328072$.

d) El llenado de una botella se considerará defectuoso si la cantidad de zumo que contiene es inferior a cierta cantidad. Determinar cuál debe ser esa cantidad si se desea que el llenado de las botellas sean considerado como defectuoso sólo en el 5 % de los casos.

Tenemos que encontrar una cantidad de zumo c tal que $P(X \leq c) = 0.05$. Nótese que entonces c es cuantil 0.05 de la distribución $N(2, 0.1^2)$. Para calcularlo, seleccionamos el menú `Distribuciones`→`Distribuciones continuas` →`Distribución normal`→`Cuantiles normales`. En la ventana emergente indicamos `Probabilidades` 0.05, `Media` 2, `Desviación típica` 0.1 y seleccionamos la opción `Cola izquierda`. Encontramos entonces que el valor de c es 1.835515 litros, que es la cantidad de zumo que deben contener las botellas para cumplir la condición indicada en el enunciado.