Práctica 4: Probabilidad. Distribuciones discretas.

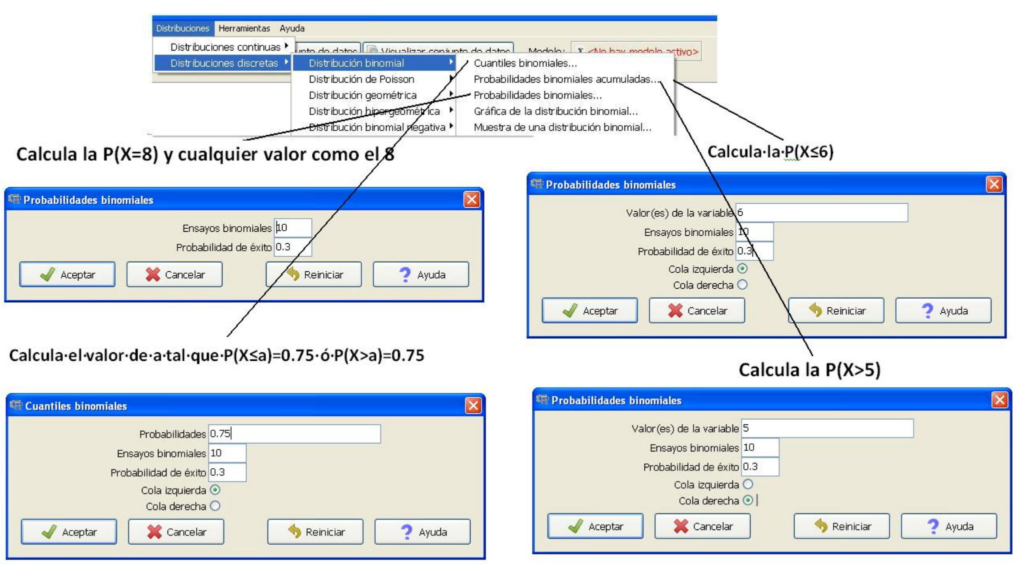
1. El menú *Distribuciones*

En este tema, trabajaremos con las distribuciones discretas estudiadas en clase. Los problemas requieren de un planteamiento previo que necesita del dominio de los conceptos acerca de estas variables. Los cálculos finales se obtienen con R-commander, su papel es el de una calculadora que realiza cálculos probabilísticos con los modelos de distribuciones con nombre propio.

El menú ***Distribuciones*** permite realizar las operaciones más habituales con los modelos de distribuciones de probabilidad más utilizados.

Ilustraremos su uso con el modelo de distribución binomial, para el resto de modelos el manejo es análogo.

1. Para la distribución binomial de parámetros $n = 10$, $p = 0{,}3$ realizamos de modo inmediato con R-commander algunos cálculos probabilísticos.



2. La opción de ***gráficas de la distribución*** proporciona tanto el gráfico de la distribución de masa de probabilidad como la de distribución.

a) Escoge los parámetros adecuados para obtener una distribución de masa asimétrica a la izquierda, otra distribución simétrica y otra asimétrica a la derecha.

b) Escoge los parámetros adecuados para obtener una distribución de masa cuya moda sea el valor 35.

Nota: La distribución geométrica en R-Commander, tiene en cuenta sólo el número de fracasos antes del primer éxito. Por ejemplo, si $X : G(p)$ y queremos calcular la probabilidad de tener éxito en el quinto intento buscaremos $P(X = 4)$ con R-Commander.

1.1. Ejercicios

1. El número de llamadas telefónicas que entran en una central telefónica se suele modelar con una distribución de Poisson. Si una central tiene un promedio 10 llamadas por hora calcula:

    a) Probabilidad de que haya exactamente 5 llamadas en una hora.

    b) Probabilidad de que haya 3 o menos llamadas en una hora.

    c) En el 80 % de las horas entrarán al menos, ¿cuántas llamadas?

    d) Probabilidad de que haya exactamente 15 llamadas en dos horas.

    e) Probabilidad de que haya 15 llamadas en 15 minutos.

2. Un fabricante de productos electrónicos de consumo espera que un 2 % de las unidades fallen durante el periodo de garantía. Se hace el seguimiento del cumplimiento de la garantía en una muestra de 500 unidades independientes.

    a) ¿Cuál es la probabilidad de que ninguna unidad falle durante el periodo de garantía?

    b) ¿Cuál es la probabilidad de que fallen más de dos unidades durante el periodo de garantía?

    c) La probabilidad de que fallen como mucho "a" unidades es 0.80, ¿cuánto es "a"?

3. En una lonja de pescado se hacen controles de modo que se analizan sucesivamente y de forma independiente pescados hasta encontrar uno con parásitos. Se conoce que la probabilidad de que un pescado este infectado con algún parásito es de 0.23.

a) ¿Qué probabilidad hay de que en un control se analicen 5 pescados?

b) En el 90 % de los controles, ¿antes de qué unidad analizada aparecerá el primer pescado infectado?

c) En qué porcentaje de controloess aparecerá la primera unidad infectada en el primer pescado observado?

4. El envío de un mensaje por teléfono móvil requiere que se envíe una llamada de aceptación a la antena más próxima. Espera una respuesta durante 0.5 segundos, si no llega, vuelve a intentarlo hasta 6 veces. Si la sexta llamada no es exitosa se genera una mensaje de saturación de red para el usuario. Calcula el porcentaje de mensajes no enviados si sabemos que la probabilidad de que la antena no responda en cada llamada es $p = 0{,}02$. ¿Cuánto debería valer p para que el porcentaje de mensajes no enviados fuera menor que 0.01?

5. Resolver el problema de control de calidad de la hoja de problemas:

   Un distribuidor recibe un lote muy grande de componentes. El lote se puede clasificar como aceptable solo si la proporción de componentes defectuosos es a lo sumo 0.1. El distribuidor sigue el siguiente protocolo para decidir si compra o no el lote. Selecciona al azar (sin devolución $\sim$ con devolución ya que p no cambia)n = 10 componentes y acepta el lote solo si el número de componentes defectuosos de la muestra es a lo sumo m = 2.

   a) Calcula la probabilidad de que el lote sea aceptado cuando la proporción real de piezas defectuosas es de 0.01, 0.05, 0.10, 0.20, 0.25, 0.50, 0.75.

   b) Sea p la proporción real de piezas defectuosas en el lote. Dibuja para distintos valores de p la probabilidad de que el lote sea aceptado. El gráfico obtenido se denomina curva característica de operación para el plan de muestreo.

   c) Repite los apartados a y b con n = 10 y m = 1 y n = 15 y m = 2. ¿ Cuál de los tres planes de muestreo te parece más adecuado para el distribuidor?

2. Simulación de realizaciones de una variable aleatoria

La simulación es la reproducción virtual de un proceso aleatorio real que consiste en simular observaciones de una variable aleatoria. Se aplica en los métodos de la ingeniería para la toma de decisiones porque permite simular una realidad dominada por un ambiente de incertidumbre.

Por ejemplo, si en una línea de producción salen de manera aleatoria elementos defectuosos, puede simularse la producción de un día y observar el resultado para tomar decisiones.

En esta sección haremos una introducción al uso de la simulación que además nos servirá para afianzar algunos resultados teóricos que serán claves en el tema de inferencia estadística.

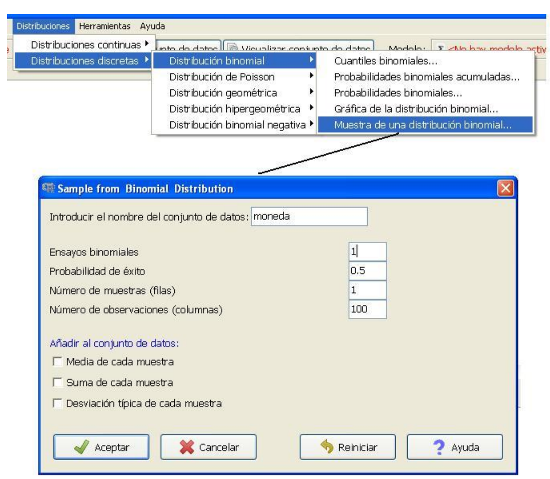
2.1. Simulación de una muestra

Se trata de simular el resultado del lanzamiento de una moneda equilibrada. Nuestra variable aleatoria X será la indicadora de cara, es decir

	X	P
Cara	1	0.5
Cruz	0	0.5

. En definitiva $X : B(n = 1, p = 0{,}5)$

1. Simulamos 100 lanzamientos de una moneda equilibrada, es decir simulamos 100 observaciones de X. En el lenguaje estadístico se dice que se simula una muestra de X de 100 observaciones. Para ello se escoge la opción *muestra de la distribución binomial* y la ventana que se abre se completa como en la figura siguiente

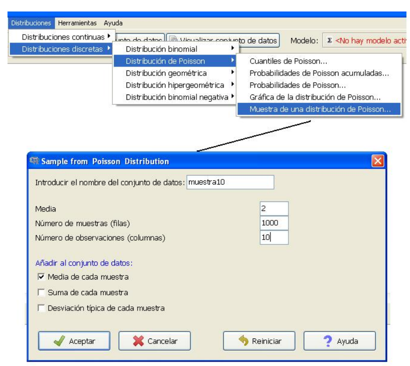


2. Observa que queda almacenado el resultado como el conjunto de datos activo que se denomina "moneda" y que puede ser guardado en un fichero de tipo .rda.

2.2. Simulación de varias muestras

Trataremos ahora con otro ejemplo. Se considera que el número de saltamontes, X, en el medio rural de un país se distribuye como una Poisson de media 2 saltamontes por metro cuadrado. A un grupo de 1000 ingenieros agrónomos repartidos por todo el país se les asigna la tarea de realizar un número n de observaciones al azar en su territorio. Así, cada uno de ellos dispone de una red de un metro cuadrado y efectúa n veces el conteo de saltamontes colocando la red en n puntos distintos elegidos al azar. Así, si por ejemplo $n = 10$, tendremos que cada ingeniero coge una muestra de 10 unidades.

1. Simula el resultado del experimento para $n = 10$. Crea un conjunto de datos denominado “muestra10 ”que recoja las observaciones así como la media de las observaciones realizadas por cada uno de los profesionales (observa que se añade al conjunto de datos como una nueva columna denominada “mean”).



Es importante observar que cada ingeniero agrónomo va a obtener una realización del vector $(X_1, X_2, \ldots, X_{10})$ que se almacena en filas (una para cada ingeniero) en el archivo de datos.

El vector $(X_1, X_2, \ldots, X_{10})$ está formado por las variables X_i: número de saltamontes en un m^2 elegido al azar. Estas variables son independientes y están idénticamente distribuidas como la variable X.

Observa que $\overline{X}_{10}$ es una variable aleatoria porque según las 10 observaciones que proporcione el azar tomará un valor u otro.

Guarda la tabla de datos “muestra10 ”en un archivo rda.

2. Repite esta simulación del experimento para $n = 50$. Crea el conjunto de datos denominado "muestra50" que recoja las observaciones así como la media de las observaciones realizadas.

2.3. Análisis de los resultados de las simulaciones

¿Qué similitudes y diferencias observas en la variable media de cada muestra (mean) cuando se tienen muestras de distintos tamaños de la misma variable ?

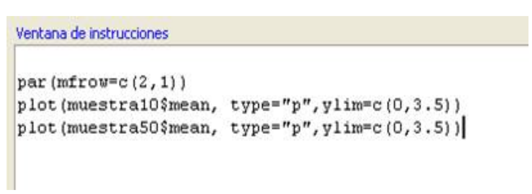
Observa que $\overline{X}_n$ es una variable aleatoria y que se varía alrededor del 2 y que esto ocurre para todos los tamaños muestrales 10 y 50. Lo que indica que la esperanza de $\overline{X}_n$ es 2, sea cual sea el tamaño de la muestra n.

Observa también que conforme el tamaño muestral (n) crece los valores de $\overline{X}_n$ se acercan más a 2. Lo que indica que la varianza de $\overline{X}_n$ decrece conforme el tamaño de la muestra (n) crece.

Visualiza estas características con gráficas secuenciales de puntos. Conviene que se vean en una misma ventana y que la escala del eje-y sea la misma para poder hacer comparaciones.

Para visualizar las gráficos en una misma ventana recuerda que la ventana de gráficos puede dividirse en dos escribiendo y ejecutando la orden **par(mfrow=c(2,1))** [2 filas y 1 columna]. Esta La división se mantendrá activa hasta que no se cierre la ventana de gráficos.

Unas vez preparada la ventana para recibir dos gráficos, se hacen los diagramas de modo consecutivo. Conviene modificar los comandos que hacen la representación añadiendo la opción *ylim=c(a,b)* para que todos presenten en el eje-y el mismo mínimo a y el mismo máximo b. En este caso es adecuado $a = 0$ y $b = 3{,}5$.
La ventana de instrucciones quedará como sigue.



2.4. Ejercicios

1. Repite los apartados 2.2 y 2.3 para el siguiente ejemplo.

Se tira una moneda hasta que sale cara y se anota X: número de tiradas.

Ten en cuenta que en R, la variable Geométrica cuenta los fracasos antes del primer éxito, como se ha hecho en clase.

2. En un supermercado se sabe que el número de clientes que llegan por hora sigue una distribución de Poisson de media 50 clientes.

   a) Imagina que en la primera hora todos los clientes que llegan se quedan en el supermercado más de una hora, ¿cuántos carros de compra debe haber en el supermercado para garantizar que al comienzo de la segunda hora haya una probabilidad de al menos 0.3 de que haya algún carrito libre?

   b) ¿ Qué probabilidad hay de que el número de clientes supere a 75

   c) Simula el comportamiento de llegada de clientes durante 200 horas y añade el promedio.

   d) ¿Es muy diferente el valor que toma $\bar{X}$ de la esperanza de la variable X ($E[X] = 50$)? ¿Por qué?

3. Lanzamos un dado hasta que sale 6 y consideramos X: número de lanzamientos.

   a) Dibuja la función de probabilidad.

   b) Si decidimos abandonar el juego si no sale un 6 para el décimo lanzamiento, ¿cuál es la probabilidad de que abandonemos el juego en una partida?

   c) Si decidimos abandonar el juego en el n-ésimo lanzamiento si para ese momento no se ha obtenido un 6, ¿cuánto debe valer n para que la probabilidad de abandonar el juego en cada partida sea menor o igual que 0.1.