OCW-UPM Estadística para Ingeniería Civil y Medioambiental

Autores: E. M. García del Toro, C. Hermoso, E. J. Huertas

LABORATORIO DE ESTADÍSTICA CON MATLAB

PRÁCTICA 4 - Cálculo de probabilidades con variables aleatorias continuas y discretas

Para obtener probabilidades con variables aleatorias discretas y contínuas, MATLAB dispone de la herramienta `disttool` , que permite graficar la función de probabilidad $P(X = k)$ (caso v.a. discreta) o la función de densidad de probabilidad $f(x)$ (v.a. continua) y las correspondientes funciones de probabilidad acumulada $F(x)$.

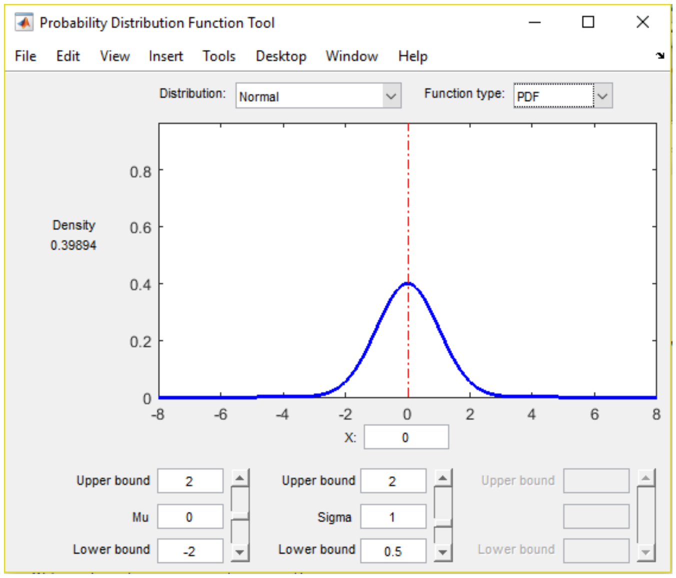

Figura 1: Ventana de la herramienta `disttool` en MATLAB

Todas las probabilidades que se calculan en esta práctica, pueden calcularse también con esta herramienta.

Parte A: Cálculo de probabilidades con variable aleatoria continua.
Función de distribución Normal (campana de Gauss).

Una *variable aleatoria continua* X puede tomar cualquier valor de entre el infinito número de valores de cierto intervalo $(a, b) \in \mathbb{R}$. Además, dicho intervalo puede ser también toda la recta real $(-\infty, +\infty) \equiv \mathbb{R}$. Por este motivo, no es posible asignar probabilidad a cada uno de los valores particulares de X, y se tiene que recurrir al concepto de integral.

Para una variable aleatoria continua, existe una función $f(x)$ llamada *función de densidad de probabilidad*, de manera que la probabilidad de obtener valores de X en un intervalo $(a, b) \in \mathbb{R}$ viene dada por:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

La distribución normal es la más importante de todas las distribuciones de probabilidad para variables aleatorias continuas. También existen muchas otras distribuciones de probabilidad para variables aleatorias continuas como la distribución Gamma, las distribuciónes de Pearson, la distribución t de Student o la distribución de Gumbel, por mencionar unas cuantas.

Para calcular probabilidades con variables aleatorias contínuas distribuídas como una $N(\mu, \sigma)$ podemos usar la herramienta `disttool`, o bien el comando:

```
normcdf(x,mu,sigma)  %devuelve la probabilidad P(X<x) de una distribucion normal con media
                      mu y desviacion estandar sigma, evaluada en X=x.
                      Los valores por defecto de mu y sigma son 0 y 1, respectivamente.
                      Representa la funcion de distribucion acumulada F(x) para los valores
```

```
                          de X~N(mu,sigma).
norminv(P,mu,sigma)  %devuelve el valor critico para la probabilidad P. Es aquel valor de la
                          variable aleatoria X que deja a su izquierda una probabilidad acumulada P.
```

Ejemplo.- Se desea estudiar los esfuerzos de compresión de una serie de probetas de cemento de una determinada mezcla, mediante la variable aleatoria contínua $X \equiv$ *"esfuerzo de compresión del cemento, medido en Newton por milímetros al cuadrado* (N/mm^2)*"*.

Se sabe que, para las probetas del estudio, X sigue una distribución normal, de media $60,14\ N/mm^2$ y desviación típica $5,02\ N/mm^2$.

a) Determine la variable aleatoria Z tipificada de la variable aleatoria X, y especifique su distribución de probabilidad.

b) ¿Cuál es la probabilidad de que un cubo modelo se rompa, si se le somete a una presión de 55 N/mm^2 o menos?

c) ¿Cuál es la probabilidad de que un cubo modelo soporte un esfuerzo de compresión entre $58,64$ y $65,91$ N/mm^2?

d) ¿Qué esfuerzo de compresión, en N/mm^2, hay que realizar para romper el 95 % de los cubos modelo sometidos a ensayo?

a) La variable aleatoria (tipificada) pedida es $Z = \frac{X-60{,}14}{5{,}02} \sim N(0;1)$

b) La variable aleatoria $X \equiv$ *"esfuerzo de compresión del cemento, medido en* N/mm^2*"* sigue una distribución $X \sim N\left(60{,}14\, N/mm^2;\ 5{,}02\, N/mm^2\right)$, luego la probabilidad de que un cubo modelo se rompa si se le somete a un esfuerzo de compresión de hasta 55 N/mm^2 es (utilizando `normcdf(x,mu,sigma)` en MATLAB)

$$P(X \leq 55) \quad = \texttt{normcdf(55,60.14,5.02)} \quad = 0{,}1529.$$

El resultado anterior varía ligeramente si usamos la tabla de la $N(0,1)$ para resolverlo, ya que ésta contiene valores aproximados. Con las tablas da un valor de 0,1539.

c) La probabilidad de que un solo cubo modelo soporte entre 58,64 y 65,91 N/mm^2 es

$$\begin{aligned} P(58{,}64 \leq X \leq 65{,}91) &= P(X \leq 65{,}91) - P(X \leq 58{,}64) \\ &= \texttt{normcdf(65.91,60.14,5.02)-normcdf(58.64,60.14,5.02)} \\ &= 0{,}87480 - 0{,}38254 = 0{,}49226 \ \ \text{(usando tablas sale 0,4928)} \end{aligned}$$

d) Hay que calcular el valor crítico a para el que la probabilidad acumulada sea de 0,95. Luego $P(X < a) = 0{,}095$. Utilizando el comando norminv tenemos

valor a, tal que $P(X \leq a) = \texttt{norminv(0.95,60.14,5.02)} = 68{,}3972\ N/mm^2$ (con tablas sale $a = 68{,}373\, N/mm^2$).

Parte B: Cálculo de probabilidades con variables aleatorias discretas.

Distribución binomial.

Dada una variable aleatoria discreta X, distribuída como una binomial de parámetros n y p, es decir $X \sim B(n,p)$, la probabilidad de que X tome uno de los $n+1$ valores discretos $k = 0, 1, 2, \ldots, n$, viene dado por la función de probabilidad discreta

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

En MATLAB podemos calcular probabilidades para esta distribución mediante los comandos:

```
binopdf(x,n,p) %devuelve el valor de la probabilidad para una distribucion binomial B(n,p)
                 con parametros n y p que hemos visto en clases de teoria.
binocdf(x,n,p) %devuelve la probabilidad acumulada desde 0 hasta el valor de x ("cola
                 izquierda o cola inferior").
binocdf(x,n,p,'upper') %devuelve la probabilidad acumulada desde x hasta n
                 ("cola derecha o cola superior").
[mu, var] = binostat(n,p)  %devuelve la media poblacional (esperanza) y la varianza de la
                 variable aleatoria discreta X.
```

Ejemplo.- Cierto parque eólico de montaña se compone de un total de 43 aerogeneradores, suficientemente separados entre sí, y que operan independientemente unos de otros. En las especificaciones de cada aerogenerador se indica que la probabilidad de ruptura frente a vientos de tipo huracanado es $p = 0{,}02$. El parque acaba de sufrir un temporal de vientos huracanados, e interesa predecir los daños ocurridos allí para enviar un equipo técnico más o menos numeroso a repararlos.

a) ¿Que distribución de probabilidad sigue la variable aleatoria $X \equiv$ *"número de aerogeneradores rotos durante el temporal, de los* 43 *operativos en el parque"?* Indicar el valor de sus correspondientes parámetros.
b) ¿Cuál es la probabilidad de que no se haya roto ningún aerogenerador?
c) ¿Cuál es la probabilidad de que se hayan roto exactamente entre 3 y 5 aerogeneradores?
d) ¿Cuál es el número esperado de aerogeneradores rotos que encontrará el equipo de reparación?

Solución:
Apartado a) Tomamos como "éxito" el que se rompa un aerogenerador durante el temporal. La ruptura (o no) de un solo aerogenerador, se puede modelar mediante un proceso de Bernoulli independiente con probabilidad de "éxito" $p = 0{,}02$ y probabilidad de "fracaso"

$$q = 1 - 0{,}02 = 0{,}98.$$

La variable aleatoria discreta X se distribuye como una **binomial** de parámetros $n = 43$ y $p = 0{,}02$, ya que con ella **contamos** el número de "éxitos" en una secuencia de $n = 43$ ensayos de Bernoulli independientes entre sí, con una probabilidad fija $p = 0{,}02$ de "éxito" entre los diferentes ensayos. Esto se escribe formalmente como

$$X \sim B(n = 43\,;\, p = 0{,}02).$$

b) La probabilidad de que no se haya roto ningún generador durante el temporal es

$$P(X = 0) \quad = \texttt{binopdf(0,43,0.02)} \quad = 0{,}41949.$$

c) La probabilidad pedida es la suma de tres probabilidades

$$\begin{aligned} P\,(3 \le X \le 5) \quad &= P\,(X = 3) + P\,(X = 4) + P\,(X = 5) \\ &= \texttt{binopdf(3,43,0.02)} + \texttt{binopdf(4,43,0.02)} + \texttt{binopdf(5,43,0.02)} \\ &= 0{,}044003 + 0{,}0089802 + 0{,}0014295 = 0{,}054413. \end{aligned}$$

d) El número pedido es el **valor esperado** de la variable aleatoria X. Lo obtenemos con el comando `[mu, var] = binostat(43,0.02)`, luego

$$\texttt{mu} = E[X] = 0{,}8600 \text{ aerogeneradores.}$$

Esto indica que es esperable que, aproximadamente, se haya roto 1 aerogenerador (en media).

Distribución de Poisson
Dada una variable aleatoria discreta X, distribuída como una binomial de parámetros n y p, es decir $X \sim B(n, p)$, la probabilidad de que X tome uno de los $n + 1$ valores discretos $k = 0, 1, 2, \ldots, n$, viene dado por la función de probabilidad discreta

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

En MATLAB podemos calcular probabilidades para esta distribución mediante los comandos

```
poisspdf(X,lambda) %devuelve el valor de la probabilidad para una distribucion de
                     Poisson(lambda) con parametro lambda que hemos visto en clases de teoria.
poisscdf(x,lambda) %devuelve la probabilidad acumulada desde 0 hasta el valor de x ("cola
                     izquierda o cola inferior").
poisscdf(x,lambda,'upper') %devuelve la probabilidad acumulada desde x hasta infinito
                     ("cola derecha o cola superior").
[mu, var] = poisstat(lambda)  %devuelve la media poblacional (esperanza) y la varianza de la
                     variable aleatoria discreta X.
```

Ejercicio. El Etna es el volcán más activo de Europa, para el que se vienen contabilizando un promedio de seis erupciones al año. Determine las probabilidades siguientes:

a) Probabilidad de que ocurran entre 1 y 3 erupciones durante los próximos 18 meses.

b) Probabilidad de que no ocurra ninguna erupción durante el próximo año.

a) Para contar el **número de erupciones durante los próximos 18 meses** definimos la variable aleatoria discreta $X \equiv$ "*número de erupciones del Etna los próximos 18 meses*". Como el promedio de número de erupciones al año (cada 12 meses) es de seis, tendremos

$$\left\{ \begin{array}{lcl} 6 \text{ erupciónes} & \rightarrow & 12 \text{ meses} \\ \lambda_a \text{ erupciones} & \leftarrow & 18 \text{ meses} \end{array} \right. , \qquad \lambda_b = \frac{18 \cdot 6}{12} = 9 \, \frac{erupciones}{18 \, meses}$$

$\lambda_a = 9 \, erupciones/18 \, meses$ y entonces $X \sim \mathcal{P}(\lambda_a = 9)$. Nos preguntan la probabilidad de que $1 \leq X \leq 3$, luego

$$\begin{aligned} P(1 \leq X \leq 3) &= P\left(X = 1\right) + P\left(X = 2\right) + P\left(X = 3\right) \\ &= \texttt{poisspdf(1,9)} + \texttt{poisspdf(2,9)} + \texttt{poisspdf(3,9)} \\ &= 0{,}0011 + 0{,}0050 + 0{,}0150 = 0{,}0211 \end{aligned}$$

b) Como ahora se refieren a otro periodo de tiempo, **tenemos que cambiar de variable aleatoria**, para contar ahora **número de erupciones del Etna en un año**. Definimos la variable aleatoria discreta $Y \equiv$ "*número de erupciones del Etna en un año*". No hace falta calcular la nueva λ_b porque es exáctamente el promedio que nos dan en el enunciado, de seis erupciones al año, luego $Y \sim \mathcal{P}(\lambda_a = 6)$, y nos preguntan por

$$P(Y = 0) \quad = \texttt{poisspdf(0,6)} = 0{,}0025$$

que es una probabilidad muy baja. Se puede dar casi por seguro que habrá erupciones durante el próximo año.