

Distribuciones estadísticas con MATLAB

M. Santos

**Modelado y Simulación de Sistemas
Facultad de Informática**

ÍNDICE:

1.- INTRODUCCIÓN A LAS FUNCIONES DE DISTRIBUCIÓN DE PROBABILIDAD	1
2.- LAS DISTRIBUCIONES EN MATLAB	2
2.1- FUNCIÓN DE DISTRIBUCIÓN DE PROBABILIDAD (PDF)	6
2.1.1- DISTRIBUCIONES CONTINUAS	6
2.1.1.1- Distribución Uniforme	6
2.1.1.2- Distribución Normal (Gausiana)	8
2.1.1.3- Distribución Exponencial (Negativa)	9
2.1.1.4- Distribución Log-Normal	10
2.1.1.5- Distribución Gamma (Erlang)	11
2.1.1.6- Distribución Beta	12
2.1.2- DISTRIBUCIONES DISCRETAS	13
2.1.2.1- Distribución Uniforme	13
2.1.2.2- Distribución Binomial	14
2.1.2.3- Distribución Binomial Negativa	16
2.1.2.4- Distribución de Poisson	17
2.2- GENERADORES DE NÚMEROS ALEATORIOS (RND)	18

1.- INTRODUCCIÓN A LAS FUNCIONES DE DISTRIBUCIÓN DE PROBABILIDAD

Variable Aleatoria

Sea S un espacio muestral, sobre el que se encuentra definida una función de probabilidad. Sea X una función de valor real definida sobre S , de manera que transforme los resultados de S en puntos sobre la recta de los reales. Se dice entonces que X es una *variable aleatoria*.

X es una función definida sobre el espacio muestral, de manera que transforma todos los posibles resultados del espacio muestral en cantidades numéricas.

Se dice que una variable aleatoria X es *discreta* si el número de valores que puede tomar es contable (ya sea finito o infinito), y si éstos pueden arreglarse en una secuencia que corresponde con los enteros positivos. En general una variable aleatoria discreta X representa los resultados de un espacio muestral en forma tal que por $P(X=x)$ se entenderá la probabilidad de que X tome el valor de x .

Se dice que una variable aleatoria X es continua si sus valores consisten en uno o más intervalos de la recta de los reales.

Función de Densidad de Probabilidad (PDF)

Al considerar los valores de una variable aleatoria es posible desarrollar un función matemática que asigne una probabilidad a cada *realización* x de la variable aleatoria X . Esta función recibe el nombre de *función de distribución (densidad) de probabilidad* de la variable aleatoria X . Se refiere a la colección de valores de la variable aleatoria y a la distribución de probabilidades entre éstos.

- La probabilidad de que un valor x ocurra es $P(x)$, expresado con un valor o un porcentaje (Discreto)
- La probabilidad de que x tenga un valor entre a y b viene dado por el área bajo la curva de $P(x)$ (Continuo)
- La probabilidad de que x tenga un valor entre $-\infty$ y $+\infty$ es 1
- La función de densidad de probabilidad es 0 o positiva

Función de Densidad Acumulativa (CDF)

La función de distribución acumulativa de la variable aleatoria X es la probabilidad de que X sea *menor o igual a un valor específico de x* .

- La probabilidad de que x tenga una valor entre $-\infty$ y a es
$$C(x) \big|_{x=a} = \int_{-\infty}^a P(x)dx$$
$$C(x) = P(X \leq x)$$
- $0 \leq C(x) \leq 1$ para cualquier x
- $C(x_i) \geq C(x_j)$ si $x_i \geq x_j$
- $P(X > x) = 1 - C(x)$

2.- LAS DISTRIBUCIONES EN MATLAB

Características clave

El Statistics Toolbox proporciona funciones que soportan:

- Modelización lineal y no lineal
- Estadística multivariante
- Estadística descriptiva
- Cálculo y ajuste de distribuciones de probabilidad
- Análisis de varianza (ANOVA)
- Verificación de hipótesis
- Estadística industrial (control de procesos estadísticos, diseño de experimentos)
- Representación gráfica estadística y gráficos interactivos

Statistics Toolbox

El Statistics Toolbox incluye una GUI interactiva que permite experimentar, describir o ajustar sus datos a una variedad de diferentes probabilidades. Por ejemplo, puede usar la GUI para representar gráficamente una función de densidad de probabilidad o una función de distribución acumulativa para investigar cómo los parámetros de distribución afectan a su posición y forma. Además, puede usar el generador de números aleatorios para simular el comportamiento asociado a distribuciones particulares. Usted puede usar entonces estos datos aleatorios para obtener modelos bajo diferentes condiciones.

El Statistics Toolbox aporta 20 distribuciones de probabilidad diferentes. Las funciones soportadas para cada distribución incluyen:

- Función de densidad de probabilidad (pdf)
- Función de distribución acumulativa (cdf)
- Inversa de la función de distribución acumulativa
- Media y varianza
- Generador de números aleatorios

Hay disponibles funciones adicionales para calcular estimaciones de parámetros e intervalos de confianza para las distribuciones guiadas por los datos, por ejemplo beta, binominal, exponencial, gamma, normal, Poisson, uniforme y Weibull.

Nota: Nos vamos a centrar en la función de densidad de probabilidad (PDF) y en los generadores de números aleatorios (RND), que son los dos tipos de funciones que usaremos en las prácticas.

Nota: teclear en MATLAB:

```
>>help stats
```

Aquí se mostrarán todas las funciones que hay para las distintas distribuciones. Para consultar la ayuda de alguna en particular, teclear help seguido de su nombre:

```
>>help normrnd
```

A continuación mostramos algunas de las funciones del Statistics Toolbox más útiles:

Distribuciones

Funciones de Densidad de Probabilidad (pdf)

betapdf - Beta density
binopdf - Binomial density
chi2pdf - Chi square density
exppdf - Exponential density
fpdf - F density
gampdf - Gamma density
geopdf - Geometric density
hygepdf - Hypergeometric density
lognpdf - Lognormal density
nbinpdf - Negative binomial density
ncfpdf - Noncentral F density
nctpdf - Noncentral t density
ncx2pdf - Noncentral Chi-square density
normpdf - Normal (Gaussian) density
pdf - Density function for a specified distribution
poisspdf - Poisson density
raylpdf - Rayleigh density
tpdf - T density
unidpdf - Discrete uniform density
unifpdf - Uniform density
weibpdf - Weibull density

Funciones de Distribución Acumulada (cdf)

betacdf - Beta cdf
binocdf - Binomial cdf
cdf - Specified cumulative distribution function
chi2cdf - Chi square cdf
expcdf - Exponential cdf
fcdf - F cdf
gamcdf - Gamma cdf
geocdf - Geometric cdf
hygecdf - Hypergeometric cdf
logncdf - Lognormal cdf
nbincdf - Negative binomial cdf
ncfcdf - Noncentral F cdf
nctcdf - Noncentral t cdf
ncx2cdf - Noncentral Chi-square cdf
normcdf - Normal (Gaussian) cdf

poisscdf - Poisson cdf
raylcdf - Rayleigh cdf
tcdf - T cdf
unidcdf - Discrete uniform cdf
unifcdf - Uniform cdf
weibcdf - Weibull cdf

Generadores de Números Aleatorios

betarnd - Beta random numbers
binornd - Binomial random numbers
chi2rnd - Chi square random numbers
exprnd - Exponential random numbers
frnd - F random numbers
gamrnd - Gamma random numbers
geornd - Geometric random numbers
hygernd - Hypergeometric random numbers
lognrnd - Lognormal random numbers
mvnrnd - Multivariate normal random numbers
mvtrnd - Multivariate t random numbers
nbirnd - Negative binomial random numbers
ncfrnd - Noncentral F random numbers
nctrnd - Noncentral t random numbers
ncx2rnd - Noncentral Chi-square random numbers
normrnd - Normal (Gaussian) random numbers
poissrnd - Poisson random numbers
random - Random numbers from specified distribution
raylrnd - Rayleigh random numbers
trnd - T random numbers
unidrnd - Discrete uniform random numbers
unifrnd - Uniform random numbers
weibrnd - Weibull random numbers

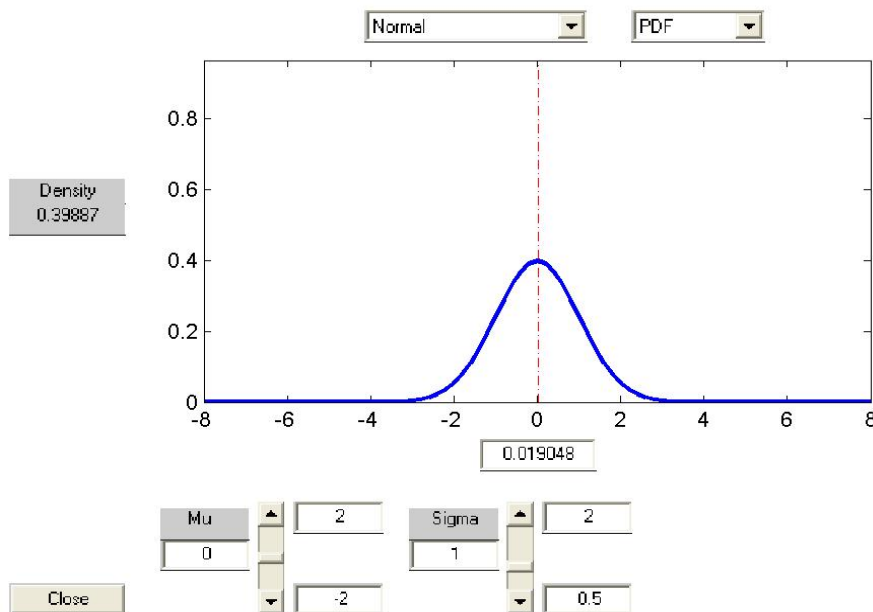
Estadísticos

betastat - Beta mean and variance
binostat - Binomial mean and variance
chi2stat - Chi square mean and variance
expstat - Exponential mean and variance
fstat - F mean and variance
gamstat - Gamma mean and variance
geostat - Geometric mean and variance
hygestat - Hypergeometric mean and variance
lognstat - Lognormal mean and variance
nbinstat - Negative binomial mean and variance
ncfstat - Noncentral F mean and variance
nctstat - Noncentral t mean and variance
ncx2stat - Noncentral Chi-square mean and variance
normstat - Normal (Gaussian) mean and variance
poisstat - Poisson mean and variance
raylstat - Rayleigh mean and variance
tstat - T mean and variance

unidstat - Discrete uniform mean and variance
unifstat - Uniform mean and variance
weibstat - Weibull mean and variance

Disttool

Es una herramienta de MATLAB que permite visualizar de forma gráfica las características de cada distribución con la posibilidad de variar sus parámetros. Las funciones que muestra son PDF y CDF. De ahí hemos obtenido las gráficas con las que ilustramos este documento.



2.1- FUNCIÓN DE DISTRIBUCIÓN DE PROBABILIDAD (PDF)

2.1.1- DISTRIBUCIONES CONTINUAS

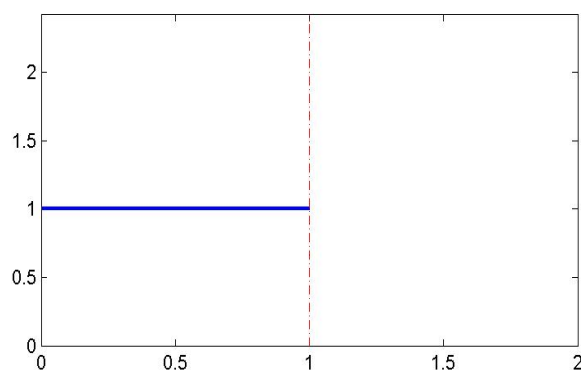
2.1.1.1- Distribución Uniforme

Sintaxis:

$Y = \text{unifpdf}(X,A,B,m,n)$

Descripción:

- Computa la función de distribución uniforme continua para el valor X y los parámetros A y B . X , A y B deben ser del mismo tamaño. El parámetro B debe ser mayor que A .
- Genera una matriz de tamaño $m \times n$, formada por números aleatorios que cumplen que su distribución sobre la recta real es de la forma indicada. (m y n son parámetros opcionales y pueden no ser necesaria su inclusión).
- El resultado Y es la probabilidad de que ocurra X en el intervalo (A,B) .
- La distribución uniforme estándar tiene $A=0$ y $B=1$.
- Ocurre en un evento donde una variable aleatoria toma valores de un intervalo finito, de manera que éstos se encuentran distribuidos igualmente sobre el intervalo. Esto es, la probabilidad de que la variable aleatoria tome un valor en cada subintervalo de igual longitud es la misma.
- Se aplica con un conocimiento muy general o poco conocimiento sobre la distribución. También para errores de redondeo en la medida, generación de números aleatorios y llegadas (cuando son independientes y el número total está determinado).



Ejemplos:

Para A y B fijados, la pdf uniforme es constante.

$X = 0.1:0.1:0.6;$

$Y = \text{unifpdf}(X)$

$Y =$

1 1 1 1 1 1

¿Qué ocurre si X no está entre A y B?

$Y = \text{unifpdf}(-1,0,1)$

$Y =$

0

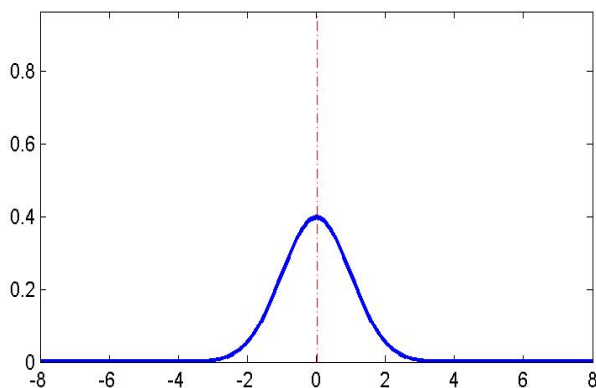
2.1.1.2- Distribución Normal (Gausiana)

Sintaxis:

$Y = \text{normpdf}(X, MU, SIGMA)$

Descripción:

- Computa la función de distribución exponencial negativa para el valor X y los parámetros MU y SIGMA. X, MU y SIGMA deben ser del mismo tamaño. El parámetro SIGMA debe ser positivo.
- La distribución normal estándar tiene $MU = 0$ y $SIGMA = 1$.
- Fluctuaciones simétricas alrededor de un valor medio (MU).
- Se aplica para describir atributos humanos o de objetos: peso, altura, etc. dentro de un grupo (variaciones en las notas de exámenes), medidas de errores angulares o lineales, generación de ruido y pequeñas perturbaciones, datos meteorológicos como temperatura y precipitación pluvial, errores de instrumentación, etc.



Ejemplos:

$X = (150:5:180);$

$Y = \text{normpdf}(X, 165, 5)$

Y =

0.0086	0.0228	0.0410	0.0499	0.0410
0.0228	0.0086			

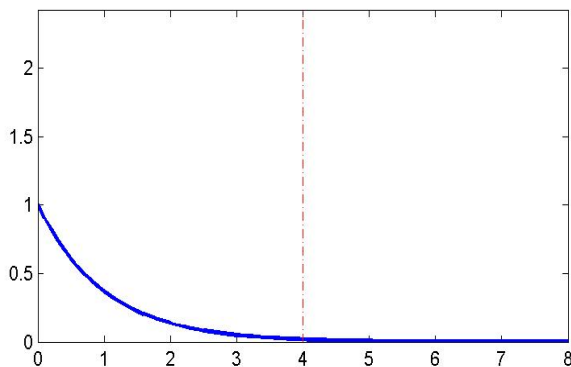
2.1.1.3- Distribución Exponencial (Negativa)

Sintaxis:

$$Y = \text{exppdf}(X, MU)$$

Descripción:

- Computa la función de distribución exponencial negativa para el valor X y el parámetro MU. X y MU deben ser del mismo tamaño. El parámetro MU debe ser positivo.
- La pdf exponencial es la pdf gamma con su primer parámetro (a) igual a 1.
- La variable aleatoria exponencial es el tiempo que transcurre hasta que se da el primer evento de Poisson. Es decir, la distribución exponencial puede modelar el lapso entre dos eventos consecutivos de Poisson que ocurren de manera independiente y a una frecuencia constante.
- Se utiliza en sucesos independientes entre sí. Es apropiada para modelar tiempos de espera cuando dicho tiempo se supone independiente del tiempo que haya transcurrido hasta ese momento. Por ejemplo, la probabilidad de que una bombilla deje de lucir en el próximo minuto es relativamente independiente del tiempo que haya estado luciendo hasta ahora. Tiempos entre roturas, tamaño de pedidos, tiempos de procesos, llegada de personas a una tienda y, en general, acciones por unidad de tiempo.



Ejemplos:

$$Y = \text{exppdf}(5, 1:5)$$

Y =

0.0067	0.0410	0.0630	0.0716	0.0736
--------	--------	--------	--------	--------

$$Y = \text{exppdf}(1:5, 1:5)$$

Y =

0.3679	0.1839	0.1226	0.0920	0.0736
--------	--------	--------	--------	--------

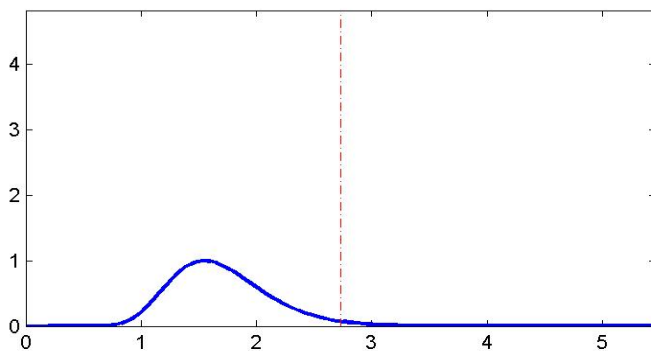
2.1.1.4- Distribución Log-Normal

Sintaxis:

$$Y = \text{lognpdf}(X, MU, SIGMA)$$

Descripción:

- Computa la función de distribución log-normal para el valor X con media MU y desviación estándar SIGMA. X, MU y SIGMA deben ser del mismo tamaño, que determina el tamaño de Y.
- Es una distribución normal asimétrica.
- Se utiliza para modelar tiempos de procesos y reparación, averías de un coche con el tiempo, población de un sitio con respecto al dinero, estancia de tiempo en un banco, etc.



Ejemplos:

$X = (1:1:10);$

$Y = \text{lognpdf}(X, 0, 1)$

$Y =$

0.3989	0.1569	0.0727	0.0382	0.0219
0.0134	0.0086	0.0057	0.0040	0.0028

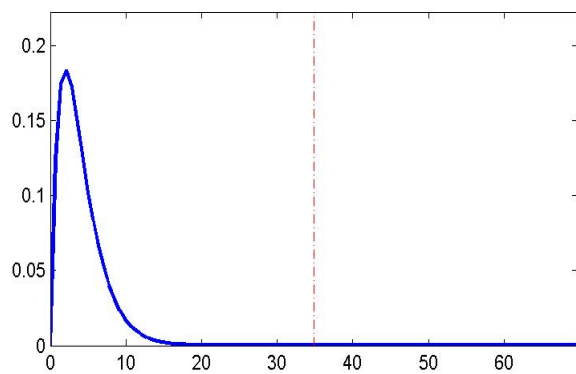
2.1.1.5- Distribución Gamma (Erlang)

Sintaxis:

$$Y = \text{gampdf}(X,A,B)$$

Descripción:

- Computa la función de distribución gamma para el valor X y los parámetros A y B. X, A y B deben ser del mismo tamaño. A y B deben ser positivos y X tiene que estar dentro del intervalo $[0,\infty)$.
- La pdf gamma es útil en los modelos de dependencia de ciclos de vida. La distribución gamma es más flexible que la exponencial. Los casos especiales de la función gamma son las funciones exponencial y chi-cuadrado.
- Se aplica para tiempos de procesos, tiempos de reparación, tiempos entre llegadas (con poca aleatoriedad), incluso ingresos familiares, edad del hombre al contraer matrimonio por primera vez, etc.



Ejemplos:

`mu = 1:5;`

`Y = gampdf(1,1,mu)`

`Y =`

0.3679	0.3033	0.2388	0.1947	0.1637
--------	--------	--------	--------	--------

`Y = exppdf(1,mu)`

`Y =`

0.3679	0.3033	0.2388	0.1947	0.1637
--------	--------	--------	--------	--------

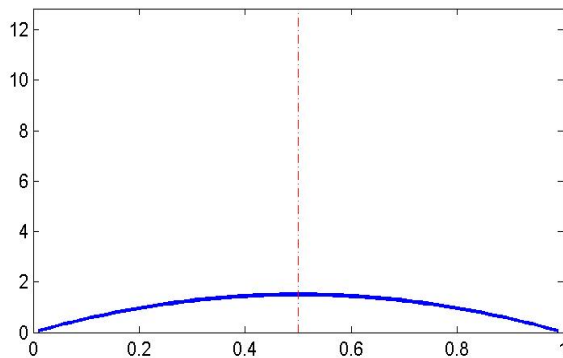
2.1.1.6- Distribución Beta

Sintaxis:

$$Y = \text{gampdf}(X,A,B)$$

Descripción:

- Computa la función de distribución gamma para el valor X y los parámetros A y B. X, A y B deben ser del mismo tamaño. A y B deben ser positivos y X tiene que estar dentro del intervalo [0,1].
- La distribución uniforme en [0,1] es un caso derivado de la distribución beta donde A=1 y B=1.
- Permite generar una gran variedad de perfiles. Se puede usar para representar la distribución de artículos defectuosos sobre un intervalo de tiempo específico, etc.



Ejemplos:

A= [0.5 1; 2 4];

Y = betapdf(0.5,A,A)

Y =

0.6366	1
1.5000	2.1875

2.1.2- DISTRIBUCIONES DISCRETAS

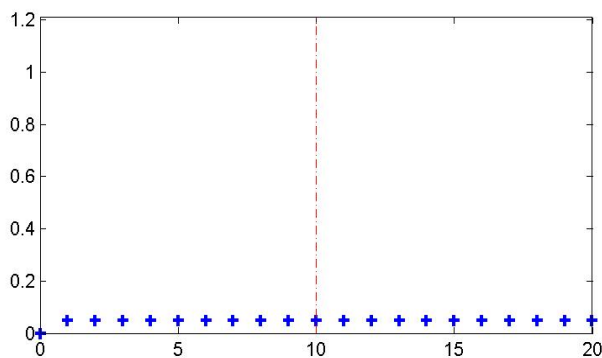
2.1.2.1- Distribución Uniforme

Sintaxis:

$$Y = \text{unidpdf}(X,N)$$

Descripción:

- Computa la función de distribución uniforme discreta para el valor X y el parámetro N . X y N deben ser del mismo tamaño y N un entero positivo.
- El resultado Y es la probabilidad de que ocurra X de entre N números, que en este caso, por ser uniforme, será la misma para cualquier X entre 1 y N .
- Se usa para generar números aleatorios de entre N . Por ejemplo podemos calcular la probabilidad de sacar un número X (del 1 al 6) al tirar un dado.



Ejemplos:

Para un N fijado, la pdf uniforme discreta es una constante.

$$Y = \text{unidpdf}(1:6,10)$$

$Y =$

0.1 0.1 0.1 0.1 0.1 0.1

Ahora fijamos x y variamos n .

$$Y = \text{unidpdf}(1:6,10)$$

$Y =$

0 0.2 0.1667 0.1429 0.1250 0.1111

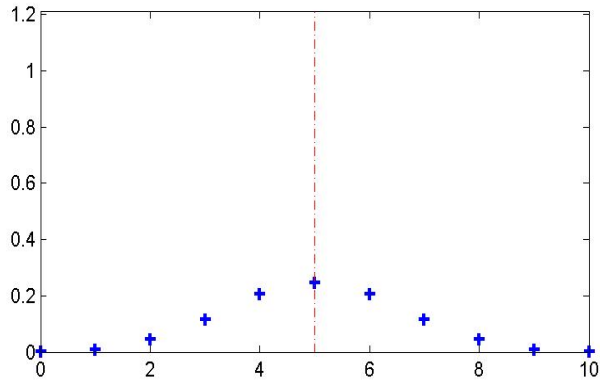
2.1.2.2- Distribución Binomial

Sintaxis:

$$Y = \text{binopdf}(X,N,P)$$

Descripción:

- Computa la función de distribución binomial para el valor X y los parámetros N y P. X, N y P deben ser del mismo tamaño. N debe ser un entero positivo y P tiene que estar en el intervalo [0,1].
- El resultado Y es la probabilidad de observar X sucesos en N pruebas independientes, donde la probabilidad de que ocurra el suceso (acierto) viene dada por el parámetro P, que permanece constante para cada prueba, y la probabilidad de que no ocurra el suceso (fracaso) es 1-P.
- Sus áreas de aplicación incluyen inspección de calidad, ventas, mercadotecnia, medicina, investigación, de opiniones y otras. Por ejemplo, un proceso de manufactura produce un determinado producto en el que algunas unidades son defectuosas; si la proporción de unidades defectuosas producidas por este proceso es constante durante un periodo razonable y, si como procedimiento de rutina, se seleccionan aleatoriamente un determinado número de unidades, entonces las proposiciones de probabilidad con respecto al número de artículos defectuosos puede hacerse mediante el empleo de la distribución binomial.



Ejemplos:

Un inspector de comercio testea al día 200 muestras de un producto. Si el 2% de ellas son defectuosas, cuál es la probabilidad de que el inspector no encuentre ninguna defectuosa ese día?

```
binopdf(0,200,0.02)
```

```
ans =
```

```
0.0176
```

Cuál es el número más probable de piezas defectuosas que se encontrará?

```
Y = binopdf([0:200],200,0.02);
```

```
[x,i] = max (Y);
```

```
i
```

```
i =
```

```
5
```

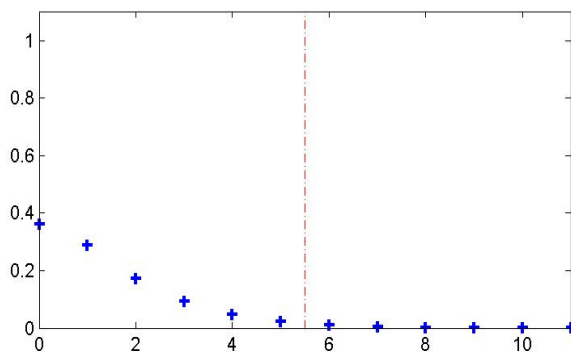
2.1.2.3- Distribución Binomial Negativa

Sintaxis:

$$Y = \text{nbinpdf}(X,R,P)$$

Descripción:

- Computa la función de distribución binomial negativa para el valor X y los parámetros R y P . X , R y P deben ser del mismo tamaño. La función de densidad es 0 a menos que X sea un entero.
- La variable aleatoria representa el número de ensayos necesarios para observar R éxitos. Los ensayos son independientes entre sí. La probabilidad de éxito en cada ensayo es constante e igual a P .
- La aplicación principal de esta distribución es una alternativa adecuada para el modelo de Poisson cuando la frecuencia de ocurrencia no es constante en el tiempo o el espacio. También se emplea para modelar las estadísticas de accidentes, datos psicológicos, compras del consumidor y otras situaciones similares en donde la frecuencia de ocurrencia entre grupos o individuos no se espera que sea la misma.



Ejemplos:

$X = (0:10);$

$Y = (X,3,0.5)$

$Y =$

0.1250	0.1875	0.1875	0.1563	0.1172	0.0820
0.0547	0.0352	0.0220	0.0134	0.0081	

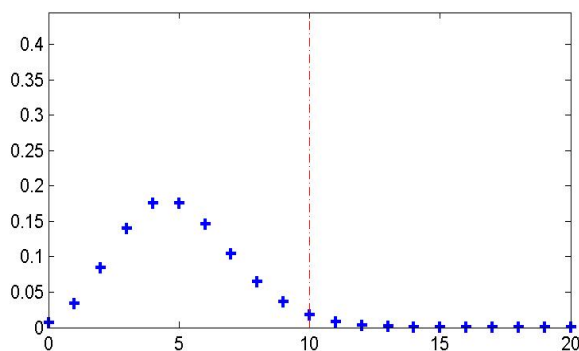
2.1.2.4- Distribución de Poisson

Sintaxis:

$$Y = \text{poisspdf}(X, \text{LAMBDA})$$

Descripción:

- Computa la función de distribución de Poisson para el valor X y el parámetro LAMBDA. X y LAMBDA deben ser del mismo tamaño y LAMBDA positivo. X puede ser cualquier entero no negativo. La función de densidad es 0 a menos que X sea un entero.
- La variable aleatoria representa el número de eventos independientes que ocurren a una velocidad constante en el tiempo o en el espacio.
- Ofrece una aproximación muy buena a la función de probabilidad binomial cuando p es pequeño y n grande.
- Es el principal modelo de probabilidad empleado para analizar problemas de líneas de espera. Se usa para estimar el número de personas que llegan a una tienda de autoservicio en un tiempo estimado, el número de defectos en piezas similares para el material, el número de bacterias en un cultivo, número de solicitudes de seguro procesadas por una compañía en un período específico, etc.



Ejemplos:

Un fabricante de discos duros para ordenadores ha observado que ocurren defectos en el proceso de fabricación aleatoriamente con una media aproximada de 2 defectos por disco de 4 Gb. Se considera este término de fallos aceptable para el proceso. Cuál es la probabilidad de que un disco duro cualquiera sea fabricado sin fallos?

Para este caso, $\text{LAMBDA} = 2$ y $X = 0$

$$P = \text{poisspdf}(0, 2)$$

P =

$$0.1353$$

2.2- GENERADORES DE NÚMEROS ALEATORIOS (RND)

Sintaxis:

R = RANDOM (<distribución>,<parámetros>)

(ó R = <distribución>rnd (<parámetros>)

Ej: R = lognrnd(...)

Descripción:

- En el apartado <distribución> se coloca el nombre de ésta entre comillas simples.
- Los parámetros corresponden a la distribución usada en cada caso y deberán cumplir las condiciones que esta exija.
- La generación de números aleatorios es útil para la simulación de eventos en los intervengan un factor aleatorio pero del que conocemos su distribución.
- Las distribuciones pueden ser: 'beta', 'Binomial', 'Chisquare', 'Exponential', 'F', 'Gamma', 'Geometric', 'Hypergeometric', 'Lognormal', 'Negative Binomial', 'Noncentral F', 'Noncentral t', 'Noncentral Chi-square', 'Normal', 'Poisson', 'Rayleigh', 'T', 'Uniform', 'Discrete Uniform', 'Weibull'.

Ejemplos:

```
» RANDOM('bino',10,0.2,1,5)
```

```
ans =
```

```
1 3 4 4 2
```

Se quiere comprobar la eficiencia de una maquina empaquetadora a la que le llegan piezas de repostería casa 5 minutos. El numero de piezas en cada remesa viene dado por una distribución uniforme con un mínimo de 5 y un máximo de 13.

```
Remesa = RANDOM('Discrete Uniform',9)+4;
```