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 3 - Variable Aleatoria Discreta. Función de probabilidad y Función de Distribución Acumulada $F(x)$

Las variables aleatorias son las herramientas matemáticas destinadas a representar los resultados de un determinado experimento aleatorio. Para ello se requiere asignar a cada elemento del espacio muestral un número. El resultado de dicha asignación es una función matemática denominada **variable aleatoria**. Así, si el experimento es lanzar una moneda y observar si sale cara o cruz, podemos asignar un 1 a la cara y un 0 a la cruz y tendremos una variable aleatoria. Las variables aleatorias se dividen en discretas, si toman valores discretos de un cierto conjunto, y continuas, si toman valores en un rango continuo. En esta práctica solo nos ocuparemos de **variables aleatorias discretas**. Las variables aleatorias solo pueden tomar valores de un conjunto conocido como soporte de la variable aleatoria. Este conjunto puede ser tanto finito como infinito, y a cada elemento del soporte se le asigna una probabilidad de que ocurra. La función que asigna una probabilidad concreta a cada posible valor que toma la variable aleatoria discreta se la conoce como **función de probabilidad** (en inglés probability mass function (pmf)). La suma de todas las probabilidades asignadas por la función de probabilidad para una determinada variable aleatoria discreta debe ser igual a 1.

Número de avenidas por año en el río Magra (Italia).

(Nota: El siguiente ejercicio está tomado del libro N. T. Kottegoda, Renzo Rosso, *Applied Statistics for Civil and Environmental Engineers*, John Wiley & Sons Ltd; 2nd Ed. (2008). ISBN-10: 1405179171 - ISBN-13: 978-1405179171)

Se considera que un río sufre una avenida cuando su caudal en un punto dado supera los 300 m^3/s. La siguiente tabla se obtuvo en la estación de medición de Calamazza, situada en el río Magra, entre las localidades de Pisa y Génova en Italia. Durante el periodo de 1939 a 1972 se tomaron los datos de los que se obtuvieron las probabilidades correspondientes a distintos números de avenidas por año que sufre el río Magra, en la zona de su cauce donde se encuentra la estación de medición de Calamazza

X:	0	1	2	3	4	5	6	7	8
$P(X = x_i)$:	0,00	0,06	0,18	0,20	0,26	0,12	0,03	0,12	0,03

Con ayuda de MATLAB cuando sea necesario, contesta a las siguientes preguntas:

1) Enuncia la variable aleatoria discreta X para la distribución anterior e identifica su soporte. ¿Es un soporte finito o infinito?

2) ¿Es esta una función de probabilidad admisible?

3) Dar un gráfico de la función de probabilidad anterior $P(X = x_i)$.

4) A la vista del gráfico del apartado anterior, ¿Dirías que esta es una función de probabilidad uniforme?

5) Calcula el número de inundaciones que se espera registrar en la estación de Calamazza para el próximo año (**valor esperado**).

6) ¿Cuál es la probabilidad de el próximo año ocurran exactamente entre 3 y 5 avenidas (incluidas ambas) en el río Magra? ¿Y de que ocurran más de tres pero menos de 5 avenidas? ¿Y tres o más, pero menos de 5 avenidas?

7) ¿Cuál es la probabilidad de el próximo año ocurran exactamente 9 avenidas en el río Magra? ¿Y más de 9 avenidas? ¿Y menos de 9 avenidas?

8) ¿Cuál es la probabilidad de el próximo año ocurra un número de avenidas estrictamente menor de 4 avenidas en el río Magra? ¿Y 4 o más avenidas?

9) Halla la Función de Distribución Acumulada o Función de Distribución $F(x)$ de la variable aleatoria X

10) Dibuja la Función de Distribución Acumulada $F(x)$ junto al gráfico de la función de probabilidad del apartado 3) ¿Qué conclusiones sacas de comparar ambas gráficas?

Solución:

1) Enuncia la variable aleatoria discreta X para la distribución anterior e identifica su soporte. ¿Es un soporte finito o infinito?

La variable aleatoria discreta puede enunciarse del siguiente modo: $X \equiv$"**Número de avenidas por año en el río Magra, a la altura de la estación de Calamazza**", y su soporte es el conjunto de valores

$$S_X = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}.$$

Es un soporte finito, ya que el conjunto de valores que puede tomar X es limitado, entre 0 y 8.

2) ¿Es esta una función de probabilidad admisible?

Como vimos en las clases teóricas, una función de probabilidades es admisible si la suma de las probabilidades de todos los posibles valores que puede tomar X es exactamente igual a 1. Para comprobarlo con MATLAB simplemente introducimos la tabla del enunciado mediante:

```
X=[0 1 2 3 4 5 6 7 8];
P=[0.00 0.06 0.18 0.20 0.26 0.12 0.03 0.12 0.03];
```

y sumamos todos los valores de `P` con el comando `sum(P)`.

3) Dar un gráfico de la función de probabilidad anterior $P(X = x_i)$.

Para dibujar la función de probabilidad ejecutamos el siguiente código MATLAB

```
figure
stem(X, P, 'filled', 'b')
% Adjust the axis limits
axis([0 9 -0.001 0.4])
% Add title and axis labels
title('Función de Probabilidad de X')
xlabel('Número de avenidas por año en Calamazza')
ylabel('Probabilidad')
```

4) A la vista del gráfico del apartado anterior, ¿Dirías que esta es una función de probabilidad uniforme?

5) Calcula el número de inundaciones que se espera registrar en la estación de Calamazza para el próximo año.

Este dato es la esperanza o valor esperado $\mu = \mathbb{E}[X]$ de la variable aleatoria X

$$\begin{aligned}
\mu &= \sum_{i=1}^{9} x_i \cdot P(X = x_i) \\
&= x_1 \cdot P(X = x_1) + x_2 \cdot P(X = x_2) + \cdots + x_9 \cdot P(X = x_9) \\
&= 0 \cdot 0{,}00 + 1 \cdot 0{,}06 + 2 \cdot 0{,}18 + 3 \cdot 0{,}20 + 4 \cdot 0{,}26 + 5 \cdot 0{,}12 + 6 \cdot 0{,}03 + 7 \cdot 0{,}12 + 8 \cdot 0{,}03 \\
&= 3{,}92 \text{ avenidas durante el próximo año.}
\end{aligned}$$

Con MATLAB podemos ejecutar el siguiente código: `Xt=X';P*Xt`

6) ¿Cuál es la probabilidad de el próximo año ocurran exactamente entre 3 y 5 avenidas (incluidas ambas) en el río Magra? ¿Y de que ocurran más de tres pero menos de 5 avenidas? ¿Y tres o más, pero menos de 5 avenidas?

Explicación en el aula.

7) ¿Cuál es la probabilidad de el próximo año ocurran exáctamente 9 avenidas en el río Magra? ¿Y más de 9 avenidas? ¿Y menos de 9 avenidas?

Explicación en el aula.

8) ¿Cuál es la probabilidad de el próximo año ocurra un número de avenidas estrictamente menor de 4 avenidas en el río Magra? ¿Y 4 o más avenidas?

Explicación en el aula.

9) Halla la Función de Distribución Acumulada o Función de Distribución $F(x)$ de la variable aleatoria X

Halla la Función de Distribución Acumulada o Función de Distribución $F(x)$ de la variable aleatoria X. Con el comando `cumsum(P)` obtenemos las probabilidades acumuladas para poder cumplimentar cada uno de los tramos de la función

$$F(x) = \begin{cases} 0{,}00 & \text{si} \quad -\infty < x < 0 \\ 0{,}00 & \text{si} \quad 0 \leq x < 1 \\ 0{,}06 & \text{si} \quad 1 \leq x < 2 \\ 0{,}24 & \text{si} \quad 2 \leq x < 3 \\ 0{,}44 & \text{si} \quad 3 \leq x < 4 \\ 0{,}70 & \text{si} \quad 4 \leq x < 5 \\ 0{,}82 & \text{si} \quad 5 \leq x < 6 \\ 0{,}85 & \text{si} \quad 6 \leq x < 7 \\ 0{,}97 & \text{si} \quad 7 \leq x < 8 \\ 1{,}00 & \text{si} \quad 8 \leq x < \infty \end{cases}$$

10) Dibuja la Función de Distribución Acumulada $F(x)$ junto al gráfico de la función de probabilidad del apartado 3) ¿Qué conclusiones sacas de comparar ambas gráficas?

```
subplot(2,1,1,'align'), stem(X, P, 'filled', 'b')
axis([0 8 -0.001 0.4])
title('Funcion de Probabilidad de X')
xlabel('Num de avenidas por anio en Calamazza')
ylabel('Probabilidad')
subplot(2,1,2,'align'), stairs(X,cumsum(P))
ylabel('Funcion de Distribucion Acumulada')
```

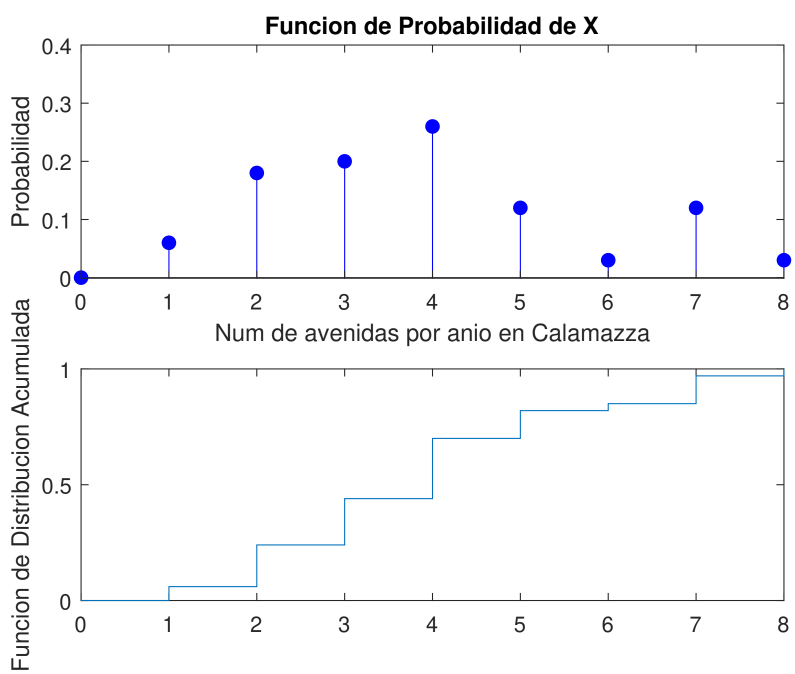

Figura 1: Funciones de Probabilidad y de Distribución Acumulada de ocurrencias de avenida X por año en la estación de aforo de Calamazza, río Magra entre Pisa y Génova, Italia.

Trabajo del alumno. La siguiente tabla muestra la probabilidad de que el número indicado de olas marinas de alta amplitud (mayores de 5 metros) golpee cierto dique en una hora.

X:	0	1	2	3	4	5	6	7
$P(X = x_i)$:	0,38	0,22	0,18	0,13	0,09	0,06	0,03	0,01

Con ayuda de MATLAB cuando sea necesario, contesta a las siguientes preguntas:

1) Enuncia la variable aleatoria discreta X para la distribución anterior e identifica su soporte. ¿Es un soporte finito o infinito?

2) ¿Es esta una función de probabilidad admisible?

3) Dar un gráfico de la función de probabilidad anterior $P(X = x_i)$.

4) A la vista del gráfico del apartado anterior, ¿Dirías que esta es una función de probabilidad uniforme?

5) Calcula el número de olas de alta amplitud que golpearán el dique durante la próxima hora.

6) ¿Cuál es la probabilidad de durante la próxima hora golpeen el dique exactamente entre 3 y 5 olas (incluídas ambas)? ¿Y de que golpeen más de tres pero menos de 5 olas? ¿Y tres o más, pero menos de 5 olas?

7) ¿Cuál es la probabilidad de durante la próxima hora golpeen exactamente 10 olas en el dique? ¿Y más de 10 olas? ¿Y menos de 10 olas?

8) ¿Cuál es la probabilidad de durante la próxima hora golpee un número de olas estrictamente menor de 4? ¿Y 4 o más olas?

9) Halla la Función de Distribución Acumulada o Función de Distribución $F(x)$ de la variable aleatoria X.

10) Dibuja la Función de Distribución Acumulada $F(x)$ junto al gráfico de la función de probabilidad del apartado 3) ¿Qué conclusiones sacas de comparar ambas gráficas?

REFERENCIAS:

N. T. Kottegoda, Renzo Rosso, Applied Statistics for Civil and Environmental Engineers, John Wiley & Sons Ltd; 2nd Ed. (2008)

ISBN-10: 1405179171 - ISBN-13: 978-1405179171