Estudio de la función de Distribución Geométrica

Aspectos básicos.Implementación en Python

María Baeza López

Juan Jesús Dóniz Labrador

Jesús Manuel Rodríguez Falcón

Grupo (3 | E)

Técnicas Experimentales. 1*er* *curso.* 2*do* *semestre*

Lenguajes y Sistemas Informáticos

Facultad de Matemáticas

Universidad de La Laguna

La Laguna, 5 de mayo de 2014

Índice general

Índice de figuras

Índice de cuadros

Capítulo 1

Motivación y objetivos

A lo largo de este cuatrimestre en la asignatura de Técnicas Experimentales hemos recurrido a la utilización de numerosas e innovadoras herramientas para la asimilación de conceptos y resolución de numerosos problemas del ámbito científico-técnico.

En el presente trabajo, se detallarán las características de las diferentes funciones matemáticas correspondientes al área de Probabilidades, realizando un mayor hincapié en la diferentes funciones de probabilidad existentes, situaciones de la vida cotidiana que necesitan de éstas para comprenderlas y analizarlas, junto a las propiedades que acompañan a dichas funciones.

Finalmente, dentro del siguiente documento nos vamos a centrar en la función Geométrica.

1.1. Problemas resueltos

1.1.1. Problema 1: Número de hijos

Un matrimonio ha decidido tener una hija, y para ello están dispuestos a tener hijos hasta que nazca una niña.Calcular el número esperado de hijos (tanto niños como niñas) que tendrá el matrimonio. Calcular la probabilidad de que la pareja anterior acabe teniendo tres hijos o más.

Solución:

Leyendo el enunciado del problema anterior podemos observar que se trata de una variable geométrica. Supongamos que la probabilidad de tener niños es la misma que la probabilidad de tener niñas. Además, sea X la siguiente variable aleatoria:

X = número de hijos varones antes de que nazca una niña

De ahí concluímos la afirmación posterior:

$$X \to Geo(p = \tfrac{1}{2}) \Leftrightarrow P[X = k] = q^{x-1} \cdot p = \tfrac{1}{2^k}$$

Sabemos que el número esperado de hijos varones que va a tener la familia es $E[X] = \frac{q}{p} = (\frac{1}{2})/(\frac{1}{2}) = 1$, de lo que es fácil deducir que el número total de hijos varones y la niña es 2.

En otro de los apartados se nos pide calcular la probabilidad de que la pareja acabe teniendo tres hijos o más. Sabemos que la probabilidad de que la pareja acabe teniendo tres o más hijos equivale a la probabilidad de que la pareja tenga dos o más hijos varones, es decir:

$$\begin{aligned} P[X \geq 2] &= 1 - P[X < 2] \\ &= 1 - P[X \leq 1] \\ &= 1 - P[X = 0] - P[X = 1] \\ &= 1 - p - qp \\ &= 1 - \frac{1}{2} - \left(\frac{1}{2}\right) \cdot \left(\frac{1}{2}\right) \\ &= \frac{1}{4} \end{aligned} \tag{1.1}$$

1.1.2. Problema 2: Tantos en baloncesto

Un jugador de baloncesto no cesa en su intento de lanzar pelotas a la canasta que se halla situada a 2 metros de altura hasta que consiga introducir una de éstas a través del aro. Si se supone que sus tiros son independientes y que la probabilidad de anotar una canasta es de 0.8,¿ cuál es la probabilidad de que el baloncestista necesite realizar dos tiros?¿ y de que sean tres tiros, cuatro tiros, cinco tiros, etc. hasta deducir la fórmula para n tiros?¿ cuál es la probabilidad de necesitar como mucho cinco tiros?

Solución:

Supongamos que tenemos la siguiente variable aleatoria:

X = número de tiros necesarios por el baloncestista hasta anotar una canasta

Ahora bien, sabemos que los valores de p y q son los siguientes:

$$p = 0{,}8;\ q = 1 - p = 1 - 0{,}8 = 0{,}2$$

Procedamos por tanto, a calcular dichas probabilidades:

- $P[X = 1] = p = 0{,}8$
- $P[X = 2] = q \cdot p = 0{,}8 \cdot 0{,}2 = 0{,}16$
- $P[X = 3] = q \cdot q \cdot p = 0{,}8 \cdot 0{,}8 \cdot 0{,}2 = 0{,}032$
- $P[X = 4] = q \cdot q \cdot q \cdot p = 0{,}8 \cdot 0{,}8 \cdot 0{,}8 \cdot 0{,}2 = 0{,}0064$

- $P[X=5] = q \cdot q \cdot q \cdot q \cdot p = 0{,}8 \cdot 0{,}8 \cdot 0{,}8 \cdot 0{,}8 \cdot 0{,}2 = 0{,}00128$

Luego, de aquí es fácil apreciar que:

$$P[X=n] = q^{n-1} \cdot p$$

Veamos ahora cuál es la probabilidad de necesitar como máximo cinco tiros para encestar en la canasta. Para ello vemos que:

$$P[X \leq 5] = P[X=1] + P[X=2] + P[X=3] + P[X=4] + P[X=5]$$

Atendiendo a los resultados del apartado anterior nos queda:

$$P[X \leq 5] = 0{,}8 + 0{,}16 + 0{,}032 + 0{,}0064 + 0{,}00128 = 0{,}99968$$

1.1.3. Problema 3: Dados

Se lanza un dado hasta que aparece el número 4. ¿ Cuál es la probabilidad de que el número de lanzamientos sean 3?

Solución:

En este ejercicio se nos pide calcular la probabilidad de realizar sólo tres lanzamientos del dado. Para ello definamos la siguiente variable aleatoria de la manera que se muestra a continuación:

X = número de lanzamientos a efectuar con el dado hasta que aparezca el valor 4 en una de sus caras

De acuerdo con el enunciado del problema sabemos que $p = \frac{1}{6}$ y, por tanto, $q = 1 - p = 1 - \frac{1}{6} = \frac{5}{6}$. De acuerdo con esto, y recurriendo a la fórmula de recurrencia obtenida en el problema anterior (*véase el problema 1.2*)

$$P[X=n] = q^{n-1} \cdot p$$

y tomando como valor n=3 podemos obtener el siguiente resultado:

$$P[X=3] = (\tfrac{5}{6})^{3-1} \cdot (\tfrac{1}{6}) = (\tfrac{5}{6})^{2} \cdot (\tfrac{1}{6}) = (\tfrac{25}{36}) \cdot (\tfrac{1}{6}) = 0{,}1157$$

1.1.4. Problema 4: Reacción frente a un medicamento

La probabilidad de que cierto análisis clínico dé una reacción positiva es 0.4. Los resultados de los análisis son independientes unos de otros ¿ Cuál es la probabilidad de que la primera reacción positiva ocurra antes del tercer análisis?

Solución:

Definamos la siguiente variable aleatoria:

X = probabilidad de que un individuo reaccione frente al medicamento

Queremos ver la probabilidad que existe de que la reacción positiva sea antes del tercer análisis.Como los valores proporcionados por el problema son $p = 0{,}4;\ q = 1 - p = 0{,}6$ nos queda:

$$P[X < 3] = P[X = 1] + P[X = 2] = (0{,}6)^{1-1} \cdot (0{,}4) + (0{,}6)^{2-1} \cdot (0{,}4) = 0{,}64$$

Capítulo 2

Fundamentos teóricos

Introducción

Para poder hablar propiamente de la distribución geométrica primero debemos introducir algunos conceptos matemáticos, referidos a la parte de probabilidades.

En primer lugar debemos contestar a la siguientes preguntas: ¿Qué es la probabilidad?, ¿Qué se entiende por función de distribución? y ¿Qué es una distribución de probabilidad discreta?. Por último, debemos profundizar sobre el objeto de nuestra investigación: la distribución geométrica.

2.1. Conocimientos previos

La probabilidad se puede definir como una función sobre una pareja (Ω, A), donde Ω es el espacio muestral [1] y A es la familia de sucesos se extiende a otros subconjuntos cumpliendo ciertas propiedades de álgebra, de la manera:

$$P : A \rightarrow \Omega$$

Esta función cumple las siguientes propiedades:

- $P(\Omega) = 1$
- Si $A_1, A_2 \in A : A_1 \cap A_2 \neq \emptyset$ entonces $P(A_1 \cup A_2) = P(A_1) + P(A_2)$

Dado un espacio probabilístico (Ω, A, P), una *variable aleatoria* ξ sobre Ω es una función real que actúa como un .ªparato de medida"sobre los sucesos elementales:

$$\begin{array}{rccc} \xi : & \Omega & \rightarrow & \Re \\ & \omega & \mapsto & \xi(\omega) \end{array} \tag{2.1}$$

Se llama función de distribución de ξ a otra función real definida como:

$$\begin{array}{rcccl} F_\xi : & \Re & \rightarrow & \Re & \\ & x & \mapsto & F_\xi(x) & = P(\{\omega \in \Omega : \xi(\omega) \leq x\}) \end{array} \tag{2.2}$$

[1] Se llama *espacio muestral* Ω a un conjunto matemático donde cada elemento representa un resultado (concreto) de un experimento.

Cumpliendo las siguientes propiedades:

- $\lim_{x \to -\infty} F(x) = 0$; $\lim_{x \to \infty} F(x) = 1$
- F no decreciente
- F continua por la derecha

Atendiendo a la forma de su función de distribución, las variables aleatorias reales pueden ser clasificadas en tres tipos:

- **Discretas:** Pueden tomar un número discreto (finito o numerable) de valores reales.
- **Continuas:** Pueden tomar todos los valores de uno o varios intervalos de la recta real, que pueden ser acotados o no acotados.
- **Mixtos:** Son combinaciones de las dos anteriores.

2.1.1. Función de probabilidad

Dada una función de distribución ξ (Véase 2.2), se llama función de probabilidad a:

$$P_\xi(k) = P(\{\omega \in \Omega : \xi(\omega) = k\}), \text{ para todo } k \in \Re$$

En nuestro caso nos interesa estudiar la función de probabilidad de la variable discreta geométrica, la cual estudiaremos en la siguiente sección.

2.2. Distribución Geométrica

La distribución geométrica es un modelo adecuado para aquellos procesos en los que se repiten pruebas consecutivamente hasta obtener el resultado deseado (éxito).

Esta distribución se puede expresar como una cualquiera de las dos distribuciones de probabilidad discreta siguientes:

- La distribución de probabilidad del número X del ensayo de Bernoulli necesaria para obtener un éxito, contenido en el conjunto $\{1, 2, 3, ...\}$
- La distribución del número $Y = X - 1$ de fallos del primer éxito,contenido en el conjunto $\{1, 2, 3, ...\}$

Atendiendo a la primera opción presenta las siguientes características:

- El proceso consta de un número no definido de pruebas o experimentales separados o separables.El proceso concluirá cuando se obtenga por primera vez el resultado deseado(éxito)
- Cada prueba puede dar dos resultados mutuamente excluyentes: A y $\neg A$(no A) [2]

[2]En teoría de conjuntos también podemos escribirlo como $\overline{A}$ o $\Omega \backslash A$

- La probabilidad de obtener un resultado A en cada prueba es p y la de obtener un resultado no A es q, siendo $p + q = 1$. [3]
- Las probabilidades p y q son constantes en todas las pruebas, por tanto, las pruebas son independientes (Si se trata de un proceso de "extracción",este se llevara a cabo con devolución del sujeto extraído)

2.2.1. Función de probabilidad de la distribución geométrica

Tal como se ha explicado con anterioridad, esta función de probabilidad se puede escribir de dos maneras posibles dependiendo de la definición de la distribución geométrica que se utilice.

Si la probabilidad de éxito en cada ensayo es p, entonces la probabilidad de que x ensayos sean necesarios para obtener un éxito es:

$$P(X = x) = (1 - p)^{x-1} \cdot p \text{ para } x = 0, 1, 2, 3, ...$$

Equivalentemente, la probabilidad de que haya x fallos antes del primer éxito es:

$$P(Y = x) = (1 - p)^{x} \cdot p \text{ para } x = 0, 1, 2, 3, ...$$

En ambos casos se corresponde con la función geométrica. En este trabajo utilizaremos la primera definición de esta función, tanto en el aspecto teórico como en el práctico.

Obtención de la función de probabilidad

En las condiciones de una distribución geométrica, podemos tomar X como una variables aleatoria, definida como:

X= El número de pruebas necesarias para obtener por primera vez un éxito o resultado A

Esta variable se distribuirá con una distribución geométrica de parámetro p, definida como:

$$X \to Geo(p)$$

Así tendremos que la variable X es el número de pruebas necesarias para la consecución del primer éxito. De esta forma las variables aleatorias toma valores enteros a partir de cero.

La función de probabilidad $P(X)$ hará corresponder a cada valor X la probabilidad de obtener el primer éxito precisamente en la X-ésima prueba. Esto es, $P(X)$ será la probabilidad del suceso obtener $X - 1$ resultados "no A" y un éxito o resultado A en la prueba número X tenieno en cuneta que todas las pruebas son independientes y que conocemos sus probabilidades, es decir, conocemos p y q. Así tendremos:

$$suceso \equiv \underbrace{\overline{A}\ y\ \overline{A}\ y\ ...\ y\ \overline{A}}_{X-1vez}\ y\ \underbrace{\overline{A}}_{1vez} = \underbrace{\overline{A} \cap \overline{A} \cap ... \cap \overline{A}}_{X-1vez} \cap \underbrace{\overline{A}}_{1vez}$$

[3] Gracias a esta característica podemos saber el valor de p(respectivamente q)a partir de su , es decir, q(respectivamente p)

dado que se trata de sucesos independientes y conocemos las probabilidades tenemos que (Utilizando la Propiedad de probabilidades Apéndice):

$$P(x) = \underbrace{P(\overline{A}) \cdot P(\overline{A}) \cdots P(\overline{A})}_{x-1vez} \cdot P(A) = q^{n-1} \cdot p \tag{2.3}$$

Luego la función de Probabilidad nos quedaría como queríamos ver:

$$P(x) = q^{n-1} \cdot p \tag{2.4}$$

Algunos autores consideran la aleatorización como "número de pruebas anteriores al primer éxito". De esta manera el conseguir el éxito a la primera sería $X = 0$. En la siguiente representación gráfica de la función de probabilidad (Veáse la Figura 2.1)de la geométrica puede apreciarse este tipo de aleatorización.

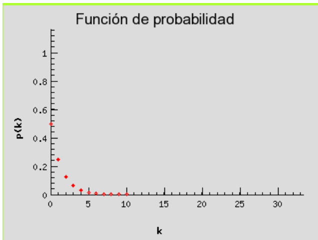

Figura 2.1: Función de probabilidad geométrica con $p = 0{,}3$

2.2.2. Función de distribución

En base de la función de probabilidad se puede expresar la función de distribución de la siguiente manera:

$$F(x) = \sum_{x=1}^{X} q^{x-1} \cdot p \tag{2.5}$$

Desarrollando la expresión tendríamos:

$$F(x) = p \cdot (1 + q + q^2 + q^3 + ... + q^{x-1}) = p \cdot \frac{1 - q^x}{1 - q} \tag{2.6}$$

de donde: $F(x) = 1 - q^x$. las gráfica análoga a la anterior:

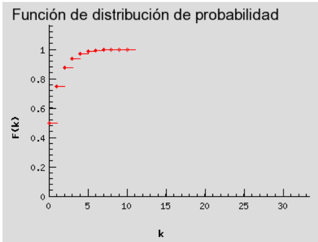

Figura 2.2: Distribución geométrica con $p = 0{,}3$

2.2.3. Propiedades de la geométrica

- **La media:** $E(X) = \frac{1}{p}$ **o** $E(Y) = \frac{1-p}{p}$

- **La varianza:**$var(X) = var(Y) = \frac{1-p}{p^2}$

- **Las funciones generatrices:**

  $G_X(s) = \frac{s \cdot p}{1-s(1-p)}$ **o también** $G_Y(s) = \frac{p}{1-s(1-p)}$, $|s| < (1-p)^{-1}$

- **La moda: es el valor de la variable que tiene asociada mayor probabilidad. Es fácil comprobar (Véase la Figura 2.1) que:**

$$P(x_i) \leq P(x = 1) \forall x_i$$

- **La mediana** M_e**: será aquel valor de la variable en el cual la función de distribución toma el valor 0.5. Así:** $F(M_e) = 1 - q^M = \frac{1}{2} \Rightarrow q^M = \frac{1}{2}$ **,por lo que,** $M_e \cdot ln(q) = ln(1) - ln(2) = -ln(2) \longrightarrow M_e = \frac{-ln(2)}{ln(q)}$

- **Función de Momentos (F.G.M):**

$$\varphi(t) = E[e^{tx}] = \sum_{x=1}^{\infty} e^{tx} \cdot p \cdot q^{x-1} = \frac{p}{q} \sum_{x=1}^{\infty} (e^t \cdot q)^x = \lim_{x \to -\infty} \frac{p}{q} \left(qe^t + (qe^t)^2 + ... + (qe^t)^x \right)$$

$$= \lim_{x \to -\infty} \frac{p}{q} \left[\frac{\overset{0}{\cancel{(qe^t)^{x+1}}} - qe^t}{qe^t - 1} \right] = \frac{p}{\cancel{q}} \left[\frac{\cancel{q}e^t}{1 - qe^t} \right] = p \left[\frac{e^t}{1 - qe^t} \right] = \frac{p}{e^{-t} - q}$$

$$= p(e^{-t} - q)^{-1}$$

Por lo que queda establecida que la F.G.M. tiene la expresión:

$$\varphi(t) = p(e^{-t} - q)^{-1}$$

Siguiendo el ejemplo de las figuras anteriores obtenemos los siguientes resultados:

	Fórmula	Valor para $P = 0{,}3$
Media	$E(Y) = \frac{1-p}{p}$	1
Varianza	$var(Y) = \frac{1-p}{p^2}$	2
Coeficiente de simetría	$G_Y(s) = \frac{p}{1-s(1-p)}$	2,1213203435596

Cuadro 2.1: Propiedades geométricas

Capítulo 3

Procedimiento experimental

Introducción:

En éste capítulo, se describirán los diversos materiales utilizados en la elaboración y ralización del algoritmo diseñados en el software Python. Además, se implementará el algoritmo de la distribución geométrica, para observar los procesos en cada paso de ejecución del progrma. Finalmente, se expondrán los resultados finales y las conclusiones que de ellos se desprendan.

3.1. Descripción de los experimentos

1. **Distribución Geometrica.**

   Análisis: Vease el capítulo 2.

   Formulación Introducción de los parámetros:

   - Probabilidad p
   - Número de casos: n

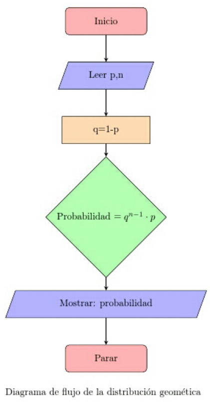

Figura 3.1: Diagrama de Flujo de la Distribución Geométrica

Algoritmo: (ver Figura 3.1).

Codifición Vease el código en en el apéndice 1.

Evalución y prueba* Va en las conclusiones.

3.2. Descripción del material

El material utilizado para los diversos programas realizados en Python:

1. **Ordenadores:**

Marca:	Modelo:	Procesador:
Acer	Aspire 5735Z	2.0 Ghz

Cuadro 3.1: Modelos de ordenadores y procesadores:

2. **Python**

3.3. Resultados obtenidos

bla, bla, etc.

Figura 3.2: Ejemplo de figura

Tiempo (± 0.001 s)	**Velocidad (± 0.1 m/s)**
1.234	67.8
2.345	78.9
3.456	89.1
4.567	91.2

Cuadro 3.2: Resultados experimentales de tiempo (s) y velocidad (m/s)

3.4. Análisis de los resultados

bla, bla, etc.

Capítulo 4

Conclusiones

bla, bla, bla, etc.

Apéndice A

Título del Apéndice 1

A.1. Algoritmo XXX

```
####################################################################################
# Fichero .py
####################################################################################
#
# AUTORES
#
# FECHA
#
# DESCRIPCION
#
####################################################################################
```

A.2. Algoritmo YYY

```
/####################################################################################
 # Fichero .h
 ####################################################################################
 #
 # AUTORES
 #
 # FECHA
 #
 # DESCRIPCION
 #
 ###################################################################################
```

Apéndice B

Título del Apéndice 2

B.1. Otro apendice: Seccion 1

`Texto`

B.2. Otro apendice: Seccion 2

`Texto`

Bibliografía

[1] Anita de Waard. A pragmatic structure for research articles. In *Proceedings of the 2nd international conference on Pragmatic web*, ICPW '07, pages 83–89, New York, NY, USA, 2007. ACM.

[2] J. Gibaldi and Modern Language Association of America. *MLA handbook for writers of research papers*. Writing guides. Reference. Modern Language Association of America, 2009.

[3] G.D. Gopen and J.A. Swan. The Science of Scientific Writing. *American Scientist*, 78(6):550–558, 1990.

[4] Leslie Lamport. *LaTeX: A Document Preparation System*. Addison–Wesley Pub. Co., Reading, MA, 1986.

[5] Coromoto León. *Diseño e implementación de lenguajes orientados al modelo PRAM*. PhD thesis, 1996.

[6] Guido Rossum. Python library reference. Technical report, Amsterdam, The Netherlands, The Netherlands, 1995.

[7] Guido Rossum. Python reference manual. Technical report, Amsterdam, The Netherlands, The Netherlands, 1995.

[8] Guido Rossum. Python tutorial. Technical report, Amsterdam, The Netherlands, The Netherlands, 1995.

[9] ACM LaTeX Style. http://www.acm.org/publications/latex_style/.