

VARIABLES ALEATORIAS DISCRETAS

Autores: Rafael García Martín (rgarciamart@uoc.edu), Francisco J. Faulín Fajardo (ffaulin@uoc.edu).

INTRODUCCIÓN

En este *math-block* titulado "**Variables aleatorias discretas**" se describen algunas de las variables aleatorias, creemos que las más aplicación tienen en el ámbito del método MonteCarlo, y la forma en que pueden ser tanto analizadas (tabular y gráficamente) como simuladas a través de Excel.

Para cada una de las v.a. presentadas se muestra: una pequeña descripción de las aplicaciones que pueden encontrarse en la literatura, la función de probabilidad y distribución, sus estadísticos principales y sus propiedades teóricas, fundamentalmente respecto a otras variables aleatorias con las que pudieran estar relacionadas.

Para cada una de ellas se presenta un método de generación de muestras aleatorias. El lector notará que se ha hecho un esfuerzo por evitar que los mecanismos de generación se basen, contrario a lo que es habitual, en código VBA.

Varias son las razones que nos han movido a ello, en primer lugar evitar la aparición, inevitable, de las macros que contuvieran dicho código y que, querámoslo o no, arrojan siempre una sombra de amenaza para la integridad de nuestros ordenadores; en segundo lugar para reivindicar la capacidad de Excel - de las funciones propias de la hoja - para realizar tareas de mediana complejidad a espaldas de un código que, aunque sumamente inteligible, implica en cualquier caso un aparato inevitable no siempre bien asumido por el usuario final.

Finalmente, la generación de v.a. a través de código VBA, o de cualquier otro lenguaje de programación, está lo suficientemente bien tratada en la literatura como para que el lector que prefiera la utilización de métodos diferentes a los aquí expuestos no tenga ninguna dificultad para encontrar información profusamente desarrollada.

OBJETIVOS

Presentar al estudiante los conceptos básicos de las variables aleatorias discretas y proporcionar una herramienta, basada en Excel, para su análisis y simulación.

RELACIÓN CON OTROS DOCUMENTOS

Este *math-block* es complementario del titulado **Variables aleatorias continuas** con el que, como es lógico, comparte muchas características. Este documento hace mención a una serie de hojas de cálculo con las que se complementa ([Binomial.xls](#) ; [BinNeg.xls](#) ; [Geometrica.xls](#) ; [Hipergeo.xls](#) ; [Poisson.xls](#) ; [UniDisc.xls](#)).

ÍNDICE

Binomial	2
Binomial Negativa	3
Geométrica	4
Hipergeométrica	6
Poisson	8
Uniforme (Discreta).....	9
Anexo 1 Hoja Patrón para las variables aleatorias discretas.	11
Anexo 2 Procedimiento genérico para la generación de variables discretas.	14
Anexo 3 Procedimiento genérico de estimación de parámetros.	19
BIBLIOGRAFÍA.....	21

Binomial

Usos.

Una v.a. Binomial representa el número de éxitos que ocurren en **n** repeticiones independientes de un ensayo de Bernoulli cuya probabilidad de éxito es **p**. Así, se distribuyen de esta forma magnitudes como el **número de piezas defectuosas** en un lote de tamaño **n** (moderado) cuando cada pieza tiene una probabilidad **p** de ser defectuosa; el **tamaño de un conjunto** si éste es aleatorio y no demasiado grande; el **número de artículos demandados en un almacén**; el **número de encuestados que están a favor de determinada cuestión** y un cuantioso etcétera.

Notación y parámetros.

La notación habitual es **X~B(n,p)**.

Probabilidad y Distribución.

La función de probabilidad es:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

La función de distribución es:

$$F(x) = \sum_{i=0}^x \binom{n}{i} p^i (1-p)^{n-i}$$

Estadísticos.

La media y varianza son (respectivamente):

$$np \quad ; \quad np(1-p)$$

Propiedades.

Si $(X_1, X_2, \dots, X_m) \sim B(n_i, p)$ entonces $(X_1 + X_2 + \dots + X_m) \sim B(n_1 + n_2 + \dots + n_m, p)$; si $X \sim B(n, p)$ entonces la variable $(n-X) \sim B(n, 1-p)$. La distribución es simétrica sólo si $p=1/2$

Generación.

Puesto que Excel cuenta con una función para la inversa de la función de distribución, la generación de variables aleatorias puede hacerse directamente por inversión utilizando la fórmula siguiente:

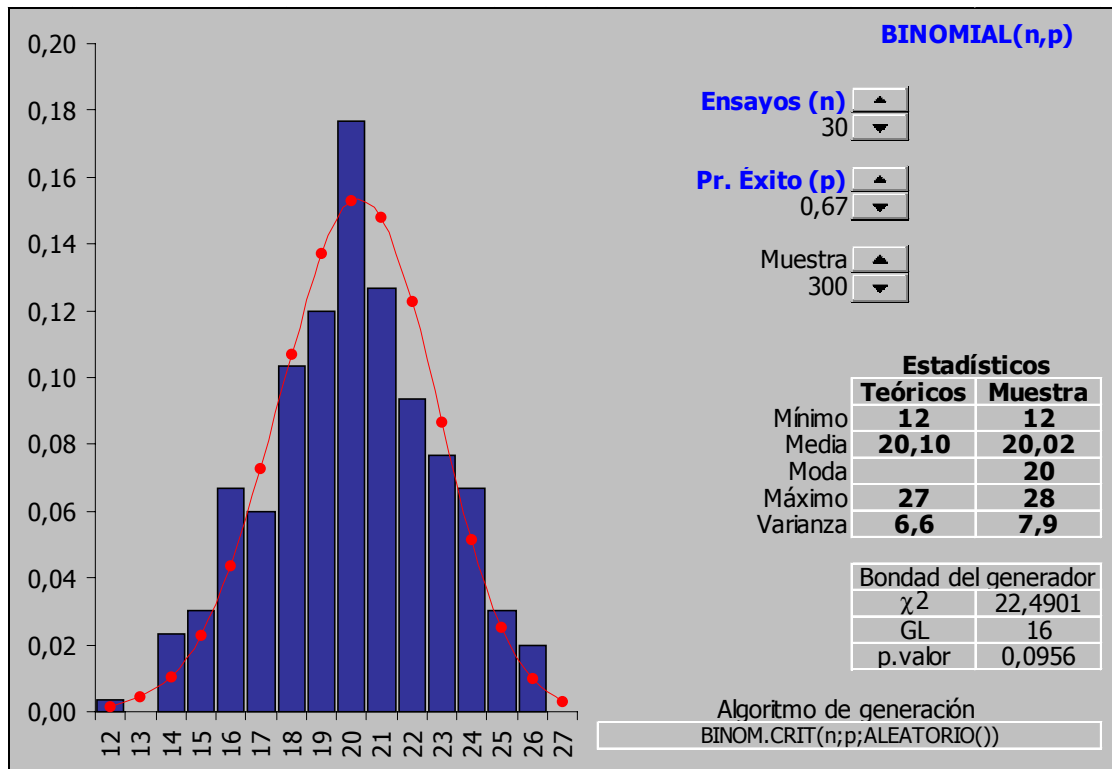
$$= \text{BINOM.CRIT}(n;p;\text{ALEATORIO}())$$

Caracterización.

Véase el Anexo 3 Procedimiento genérico de estimación de parámetros.

Hoja de cálculo.

El fichero [Binomial.xls](#) es una plantilla para la generación y análisis de la distribución Binomial en Excel así como para la estimación de parámetros a partir de una muestra aleatoria. Su aspecto es el siguiente:



Binomial Negativa

Usos.

Una v.a. Binomial negativa representa el número de fracasos que ocurren hasta obtener el n -ésimo éxito en la realización de ensayos de Bernoulli con probabilidad p de éxito. Así, el **número de artículos examinados de un lote hasta que aparece el n -ésimo defectuoso**; el **número de candidatos a entrevistar** cuando se quiere formar un equipo de n personas idóneas para un puesto de trabajo; el **número de melocotones que un cliente exigente manipula antes de conseguir un kilo de ellos que satisfagan sus criterios**; etc.

Notación y parámetros.

La notación habitual es X -NegBin(n,p) o, a veces, $BN(n,p)$.

Probabilidad y Distribución.

La función de probabilidad es:

$$p(x) = \binom{n+x-1}{x} p^x (1-p)^x$$

La función de distribución es:

$$F(x) = \sum_{i=0}^{x-1} \binom{n+i-1}{i} p^n (1-p)^i$$

Estadísticos.

La media y varianza son respectivamente.

$$\frac{n(1-p)}{p} ; \frac{n(1-p)}{p^2}$$

Propiedades.

Si $(X_1, X_2, \dots, X_m) \sim \text{BN}(n_i)$ entonces $(X_1 + X_2 + \dots + X_m) \sim \text{BN}(n_1 + n_2 + \dots + n_m)$. También es conocida como distribución de **Pascal** o distribución de **Polya**. Se verifica que **BN(1,p) \equiv Geom(p)**.

Generación.

Excel cuenta con una función para la distribución y probabilidad de la Binomial Negativa aunque no con la inversa de la distribución. No cuenta tampoco con la posibilidad de obtener muestras aleatorias a partir del módulo de Análisis de Datos + Generación de números aleatorios.

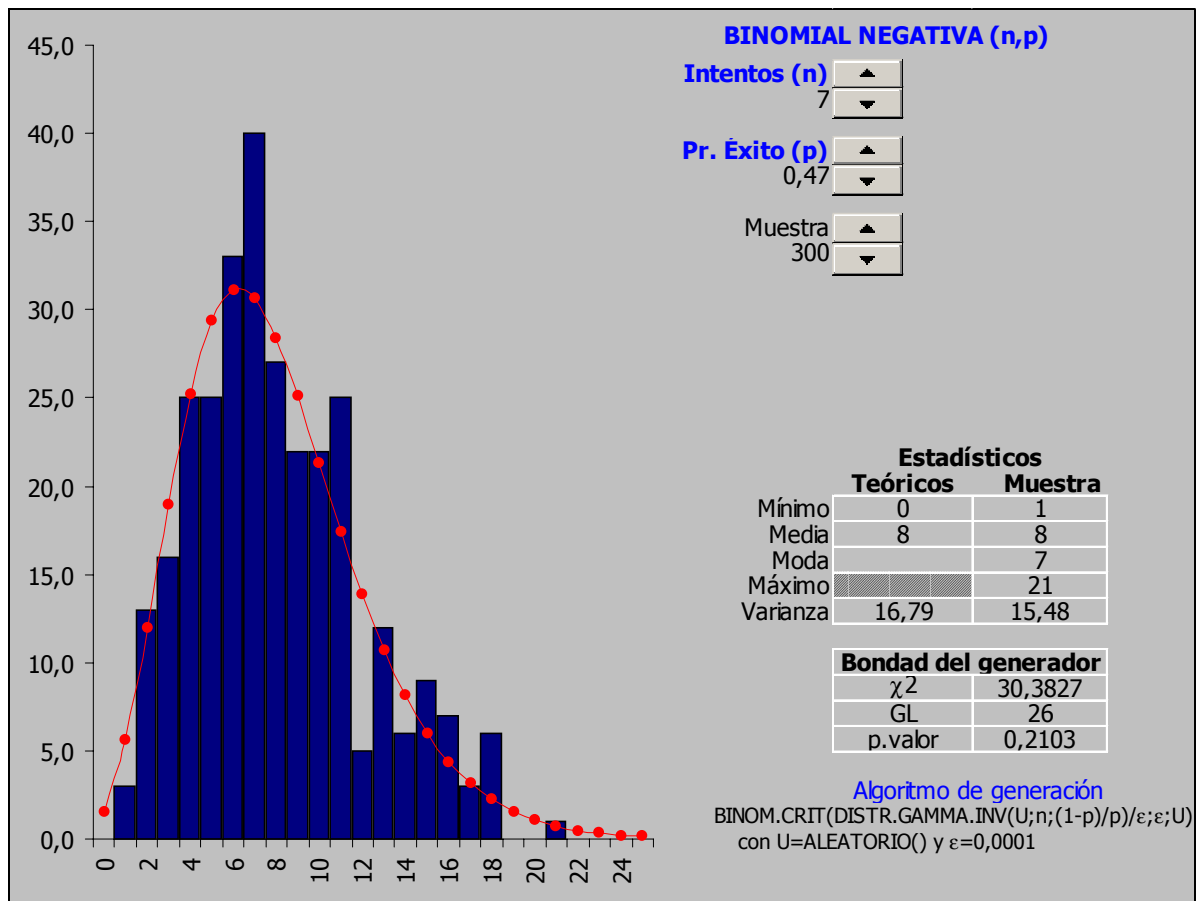
En cualquier caso es posible obtener números que se distribuyan según una esta distribución utilizando la fórmula siguiente¹:

$$\text{BINOM.CRIT}(\text{DISTR.GAMMA.INV}(U;n;(1-p)/p)/\epsilon;\epsilon;U)$$

siendo ϵ un número suficientemente pequeño (obtendremos buenos resultados con $\epsilon = 0,0001$) y U la Uniforme (0;1), es decir $U = \text{ALEATORIO}()$.

Hoja de cálculo.

El fichero [BinNeg.xls](#) es una plantilla para la generación y análisis de esta distribución:



Geométrica

Usos.

Una v.a. Geométrica representa el número de fracasos que ocurren hasta obtener el primer éxito en la realización de ensayos de Bernoulli con probabilidad **p** de éxito. Así, el **número de artículos examinados de un lote hasta que aparece el primer defectuoso**; el **número de**

¹ Por composición siguiendo una conocida propiedad de la distribución BN, véase por ejemplo [3].

candidatos a entrevistar cuando se quiere encontrar una única persona idónea para un puesto de trabajo; el **número de melones que un cliente exigente manosea antes de conseguir aquél que satisface sus criterios**, etc.

Notación y parámetros.

La notación habitual es **X-Geom(p)** o, a veces, **G(p)**.

Probabilidad y Distribución.

La función de probabilidad es:

$$p(\mathbf{x}) = p(1-p)^{\mathbf{x}}$$

La función de distribución es:

$$F(\mathbf{x}) = 1 - (1-p)^{\mathbf{x}+1}$$

Estadísticos.

La media y varianza son respectivamente.

$$\frac{(1-p)}{p} ; \frac{(1-p)}{p^2}$$

Propiedades.

La primera propiedad es evidente: se trata de una particularización de la binomial negativa, es decir, se verifica que **BN(1,p) ≡ Geom(p)**. Si $(X_1, X_2, \dots, X_m) \sim G(p)$ entonces $(X_1 + X_2 + \dots + X_m) \sim BN(m, p)$.

Es el equivalente discreto de la Exponencial en el sentido de que es la única distribución discreta que **"no guarda memoria"** ya que el número de fallos ocurridos hasta un instante dado no modifica la probabilidad de que el próximo intento sea un éxito.

Generación.

Excel no cuenta con una función para la distribución y probabilidad de la distribución Geométrica, sin embargo es fácil generar muestras aleatorias por inversión de la función de Distribución utilizando la fórmula siguiente

$$\text{REDONDEAR.MENOS(LN(ALEATORIO())/LN(1-p);0)$$

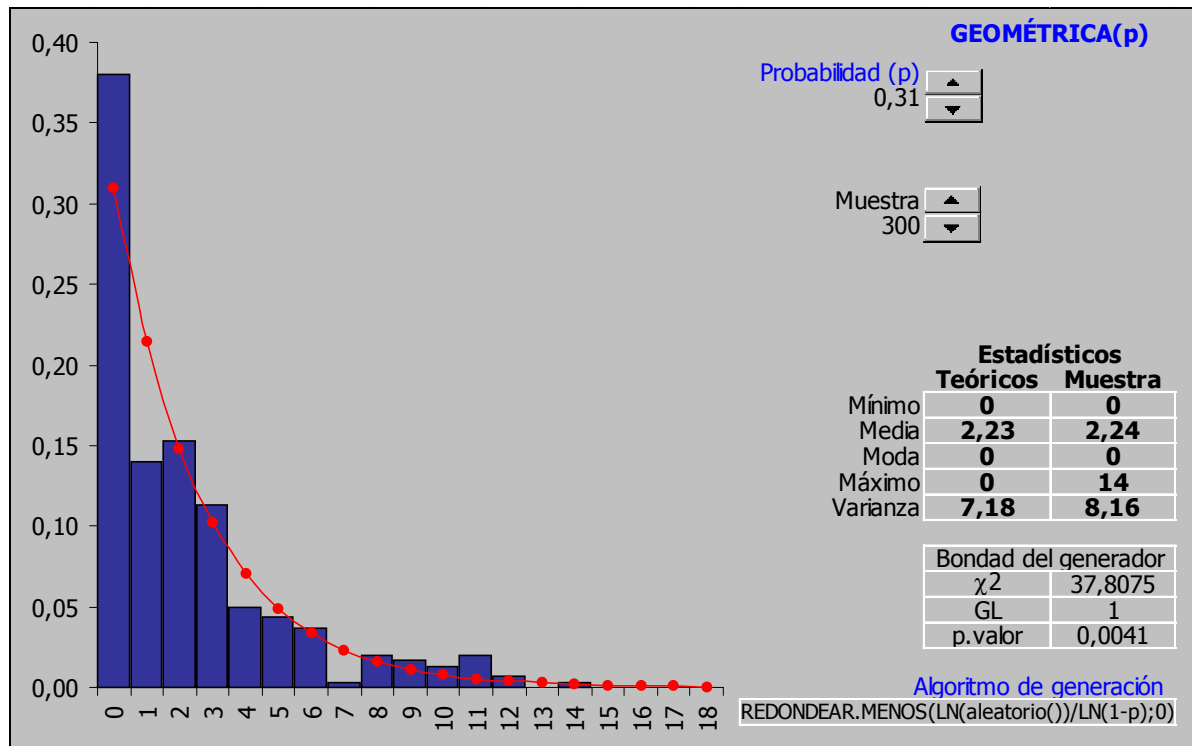
Caracterización.

Es trivial ya que se verifica que:

$$\hat{p} = \frac{1}{\bar{\mathbf{x}}_{(n)} + 1}$$

Hoja de cálculo.

El fichero [Geometrica.xls](#) es una plantilla para la generación y análisis de esta distribución:



Hipergeométrica

Usos.

Una v.a. Hipergeométrica representa el número de éxitos que ocurrirán cuando de una población en la que hay **N** objetos cuya elección se considera (arbitrariamente) un éxito y **M-N** objetos fracaso, se extrae una muestra, sin repetición, de tamaño **n**. Es importante notar que el muestreo se hace sin repetición, es decir sin devolver los objetos al seno de la población antes de cada ensayo, ya que esta característica es la única que diferencia esta distribución de la distribución binomial.

Se distribuyen según una Hipergeométrica magnitudes tales como el número de hombres (o de mujeres) que incluye una selección al azar de un grupo en el que ambos géneros están presentes, el número de temas estudiados por un opositor que ha decidido estudiar sólo unos cuantos del temario de su oposición cuando el examen consta de varios temas, etc.

Notación y parámetros.

La notación habitual es **X~HiperGeom(n,N,M)** o también **X~H(n,N,M)**. Todos los parámetros deben ser lógicamente positivos y representan: **n** el tamaño de la muestra extraída; **N** el número de éxitos que contiene la población; **M** el número total de elementos de la población.

Probabilidad y Distribución.

La función de probabilidad es:

$$p(x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$$

La función de distribución es:

$$F(x) = \frac{1}{\binom{N}{n}} \sum_{i=0}^{i=x} \binom{M}{i} \binom{N-M}{n-i}$$

Estadísticos.

La media y varianza son:

$$\frac{nM}{N} ; \left(\frac{N-n}{N-1} \right) \left(\frac{nM}{N} \right) \left(1 - \frac{M}{N} \right)$$

Propiedades.

Es evidente que ha de verificarse que: $\text{Max}(0, n - N + M) \leq X \leq \text{Min}(M, n)$

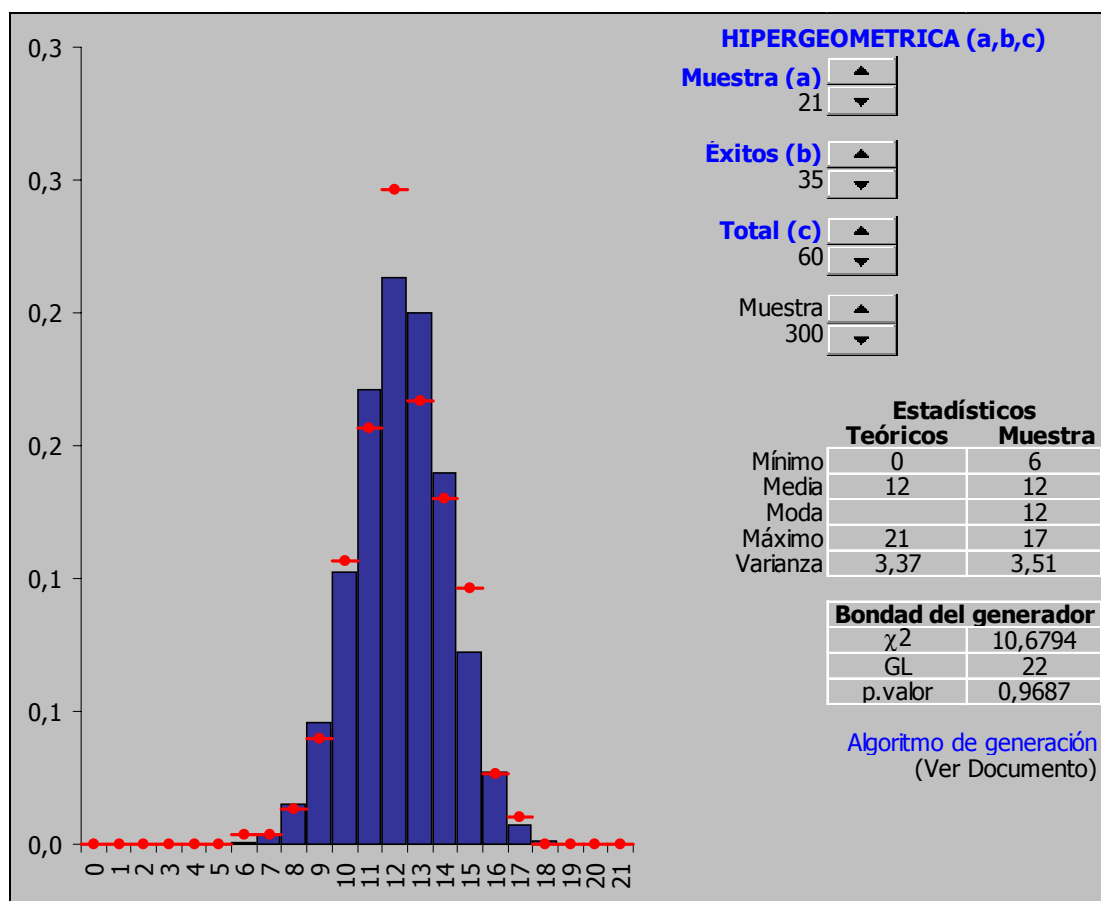
Generación.

Excel cuenta con una función para la distribución y probabilidad, no cuenta sin embargo, con la posibilidad de obtener muestras aleatorias (Análisis de Datos). Sin embargo, es posible obtener números que se distribuyan según esta distribución a través de la fórmula matricial siguiente²:

COINCIDIR(ALEATORIO());
1-PROBABILIDAD(FILA(INDIRECTO("A1:A"&n+1))-1;
DISTR.HIPERGEOMFILA(INDIRECTO("A1:A"&n+1))-1;n;N;M)
FILA(INDIRECTO("A1:A"&n+1))-1;n+1;1)
-1)

Hoja de cálculo.

El fichero [Hipergeo.xls](#) es una plantilla para la generación y análisis de esta distribución:



² Véase el procedimiento genérico expuesto en el Anexo II de este documento.

Poisson

Usos.

Una v.a. de Poisson es en realidad una v.a. Binomial llevada al límite, es decir cuando $n \rightarrow \infty$ (aunque basta con que sea suficientemente grande) y $p \rightarrow 0$ (aunque basta con que sea suficientemente pequeño).

En general cualquier **suceso "raro"** puede ser perfectamente modelizado por un v.a. de Poisson, ejemplos típicos son el **número de remaches defectuosos en un avión** (porque un avión puede llegar a tener varios millones de ellos y al ser un mecanismo tan simple es realmente difícil que sea defectuoso); el **número de erratas en un libro** (que contiene un gran número de palabras que difícilmente están mal escritas); el **número de llegadas a un servicio** (o de llamadas a un *call-center*) si la distribución entre los tiempos es exponencial; el **número de accidentes laborales en un mes en una gran empresa**; el **número de personas que entran en un supermercado en un minuto**; el **número de personas residentes en una gran ciudad que en un día sufren un infarto**; etc.

Notación y parámetros.

La notación habitual es **X~Poisson(λ)**. El único parámetro debe ser positivo $\lambda > 0$.

Probabilidad y Distribución.

La función de probabilidad es:

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

La función de distribución es:

$$F(x) = e^{-\lambda} \sum_{i=0}^x \frac{\lambda^i}{i!}$$

Estadísticos.

La media y varianza coinciden en el único parámetro λ .

Propiedades.

Si $(X_1, X_2, \dots, X_m) \sim \text{Poisson}(\lambda_i)$ entonces $(X_1 + X_2 + \dots + X_m) \sim \text{Poisson}(\lambda_1 + \lambda_2 + \dots + \lambda_m)$.

Generación.

Excel cuenta con una función para la distribución y probabilidad de Poisson, cuenta también con la posibilidad de obtener muestras aleatorias así distribuidas (Herramientas + Análisis de Datos + Generación de números aleatorios). En cualquier caso es posible obtener números aleatorios que se distribuyan según una Poisson de parámetro λ , utilizando la fórmula siguiente:

BINOM.CRIT($\lambda/0,001;0,001;ALEATORIO()$)

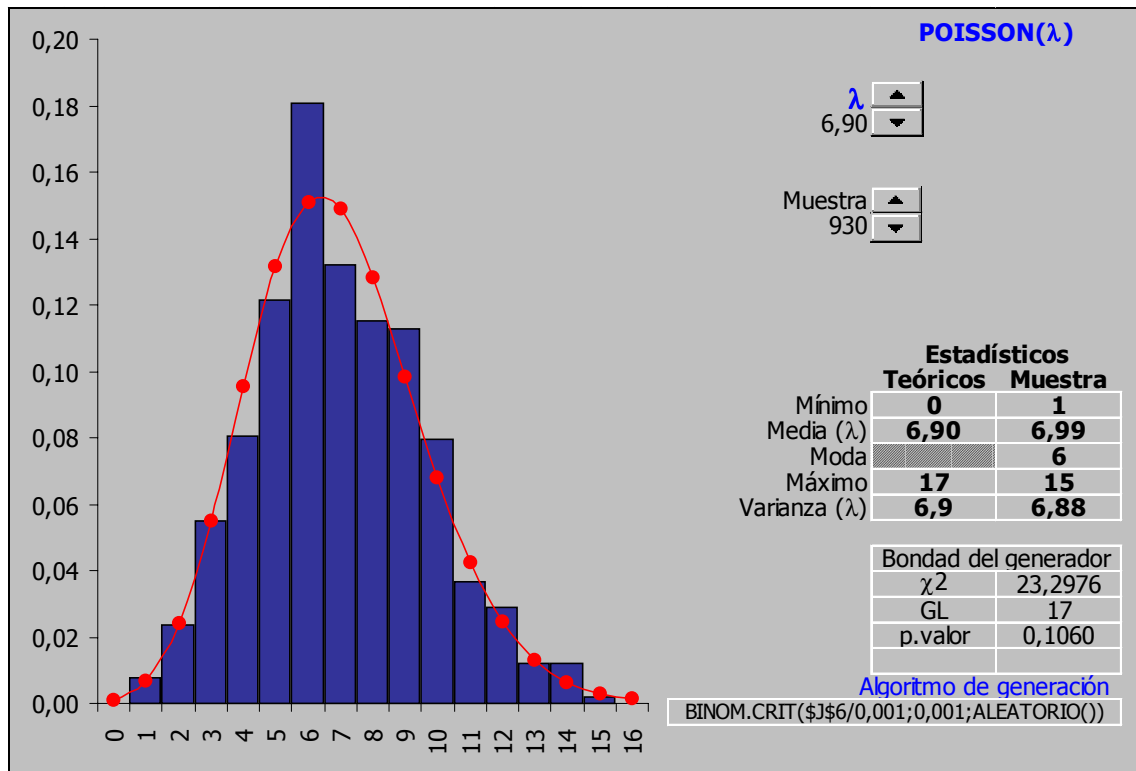
Caracterización.

El parámetro λ puede ser estimado fácilmente de la forma siguiente:

$$\hat{\lambda} = \bar{x}_{(n)}$$

Hoja de cálculo.

El fichero [Poisson.xls](#) es una plantilla para la generación y análisis de esta distribución:



Uniforme (Discreta)

Usos.

Esta v.a. es el equivalente discreto de la de mismo nombre dentro de las distribuciones continuas. Se utiliza cuando **un conjunto de posibles resultados es igualmente probable**: el número de veces que aparecerá cada una de las caras de un dado regular; el dígito final de los números premiados en la lotería, etc.

Notación y parámetros.

La notación habitual es **X-UD(a,b)**. Se verifica que $a \leq X \leq b$ y $a < b$.

Probabilidad y Distribución.

La función de probabilidad es:

$$p(x) = \frac{1}{b-a+1} \quad x = \{a, a+1, a+2, \dots, b-2, b-1, b\}$$

La función de distribución es:

$$F(x) = \frac{x-a+1}{b-a+1}$$

Estadísticos.

La media y varianza son:

$$\frac{a+b}{2} ; \frac{(a-b+1)^2 - 1}{12}$$

Generación.

Excel cuenta con una función directa para generar muestras aleatorias así distribuidas

ALEATORIO.ENTRE(a;b)

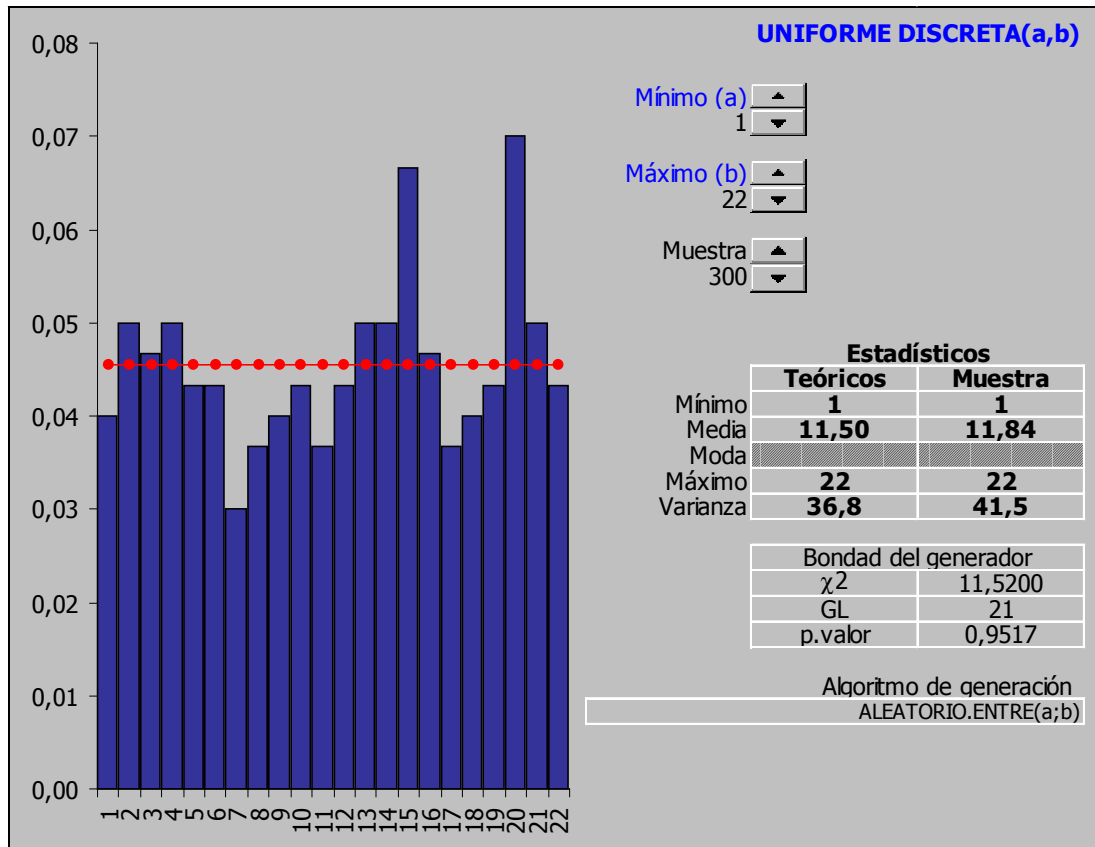
Caracterización.

Los parámetros pueden ser estimados fácilmente de la forma siguiente:

$$\hat{a} = \min\{X_{(n)}\} \quad ; \quad \hat{b} = \max\{X_{(n)}\}$$

Hoja de cálculo.

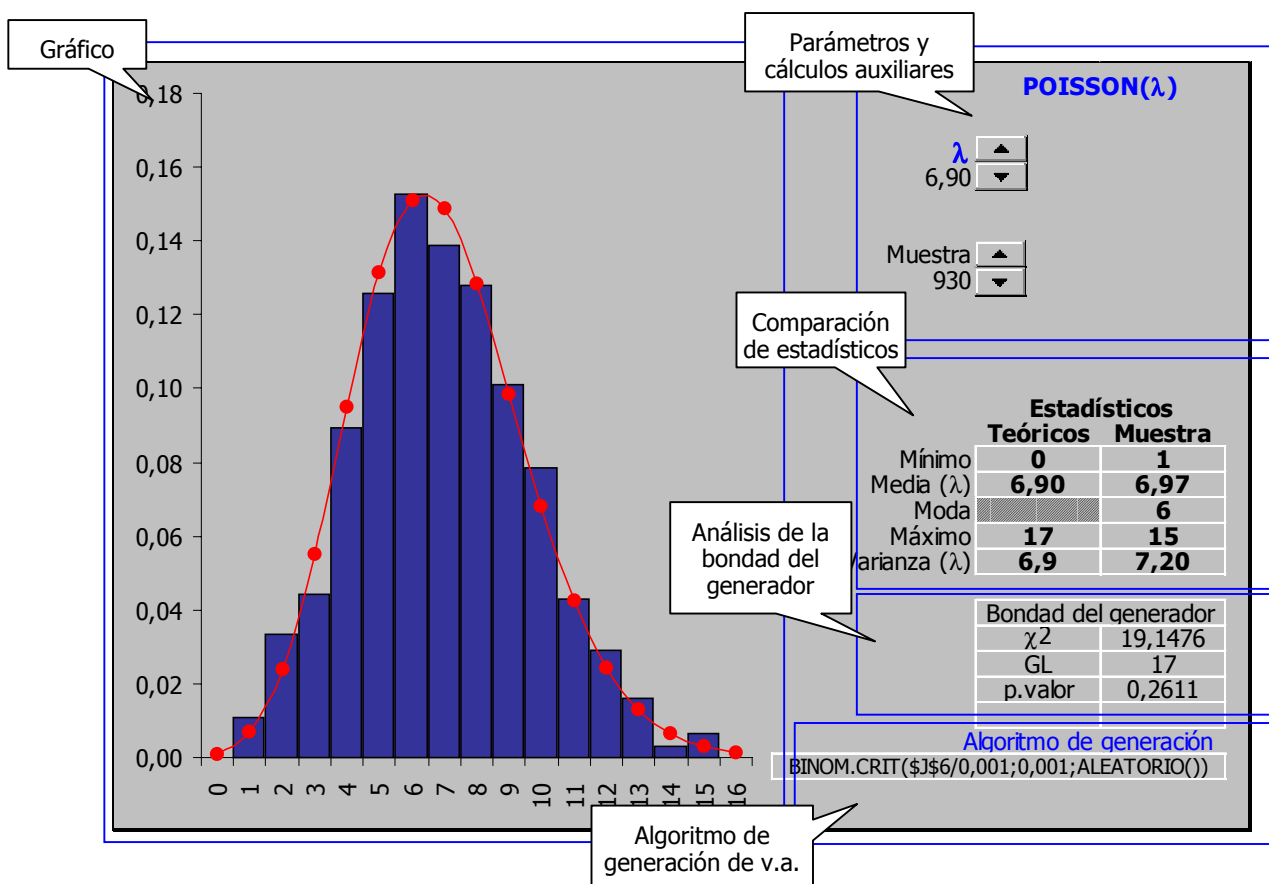
El fichero [UniDisc.xls](#) es una plantilla para la generación y análisis de esta distribución:



Anexo 1 Hoja Patrón para las variables aleatorias discretas.

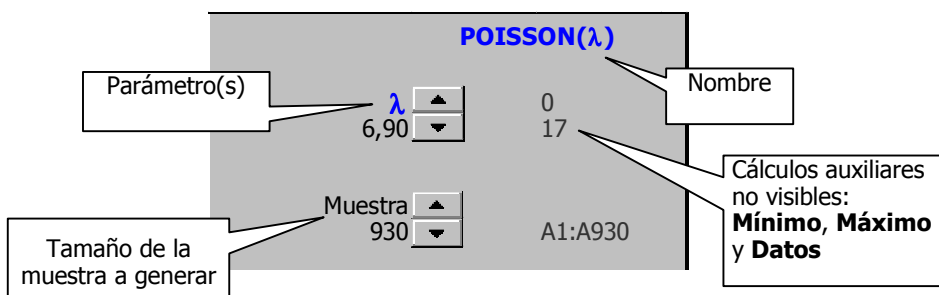
Cada una de las hojas de cálculo dedicadas al análisis de una distribución está dividida en seis partes. Cinco de ellas son externamente visibles y se han utilizado para presentar la información gráfica y numérica que describe la distribución de que se trate para la configuración de parámetros elegida por el usuario. La sexta, que permanece oculta bajo las anteriores, está dedicada a recoger la muestra aleatoria generada bajo el algoritmo propuesto y la tabulación de ésta junto con la distribución teórica, tabulación sobre la que se basa la descripción gráfica visible de la distribución.

A la vista del usuario aparecen las zonas siguientes: parámetros y cálculos auxiliares, gráfico, comparación de los estadísticos teóricos con los muestrales y la descripción del algoritmo, o algoritmos, de generación de variables aleatorias.



Parámetros y Auxiliares.

En esta zona se indica el nombre común, no abreviado, de la variable aleatoria, sus parámetros, así como un control que permite al usuario variar tanto el valor de los parámetros como el tamaño de la muestra aleatoria que se generará.



Muestra	Tamaño de la muestra simulada	Se controla por formulario, el valor máximo es 300 que puede ser modificado en las propiedades del selector.
Mínimo	Valor mínimo teórico de los datos.	En realidad se trata del valor crítico para el 1% o el 5%. Calculado utilizando bien el algoritmo de generación con un U forzado al valor correspondiente (0,01 o 0,99), bien utilizando la función de distribución inversa cuando existe para esos mismo valores
Máximo	Análogo al mínimo	Id al anterior.
Datos	Dirección en la que se encuentra la muestra simulada	Colocaremos la muestra siempre en la primera columna de manera que su dirección será ="A1:A" & Muestra

Comparación de los estadísticos.

Se calculan tanto los estadísticos teóricos de la distribución para los parámetros elegidos como los observados en la muestra generada. Esta comparación se hace únicamente a efectos informativos y no tiene ningún valor de contraste de hipótesis.

Es necesario adecuar para cada v.a. los estadísticos teóricos empleados, los muestrales se construyen con las funciones de Excel aplicadas al rango de datos simulados.

		Estadísticos	
		Teóricos	Muestra
Estadísticos muestrales y teóricos	Mínimo	-7,00	-116,81
	Media	0,00	3,88
	Máximo	7,00	1053,49
	Varianza		

Análisis de la bondad del generador.

Se realiza una prueba de bondad del ajuste de la muestra obtenida, a través del algoritmo de generación propuesto, a la distribución teórica esperada. Los resultados que se muestran son: valor del estadístico (χ^2), grados de libertad (GL), probabilidad asociada a la hipótesis nula (p.valor) de ajuste a la distribución especificada.

Resultados de la prueba de bondad del ajuste de la muestra obtenida a la distribución teórica		Bondad del generador	
		χ^2	19,1476
		GL	17
		p.valor	0,2611

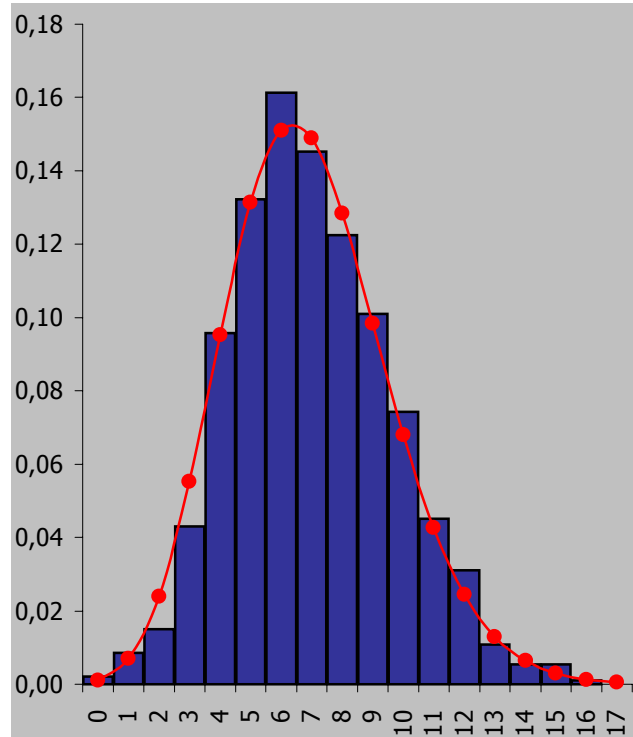
Algoritmo de generación de la v.a.

Se describe la sintaxis del algoritmo empleado en la generación de la muestra aleatoria en términos de las funciones de Excel.

Algoritmo de generación
BINOM.CRIT(\$J\$6/0,001;0,001;ALEATORIO())

Gráficos

Función de probabilidad de la v.a. e histograma de los datos simulados. Los rangos son **X** vs **Den** para la probabilidad de la v.a. y **X** vs **f1_s** para el histograma según los nombres descritos en el siguiente epígrafe.



Tabla

Ocultas tras los gráficos se encuentran la tabla en la que se realizan los cálculos que permiten la representación gráfica tanto de la variable aleatoria teórica como de la muestra simulada.

La descripción de los epígrafes de la tabla y del contenido a los que éstos se refieren es la siguiente:

X	Rango de la v.a.	El primer valor es Mínimo + Delta, los demás se calculan como $L(-1)C + \text{Delta}$
Den	Probabilidad de la v.a.	Función de probabilidad teórica evaluada en cada punto de X. Se utilizan las funciones de Excel cuando éstas existen o bien se construye la fórmula manualmente
ni	Frecuencia absoluta de la muestra en función de los valores de X	Número de observaciones de la muestra en cada clase de X. Es matricial . = $\{SUMA(SI(INDIRECTO(Indirecto)=X;1;0))\}$
f1_s	Probabilidad de la muestra	Para cada clase, se calcula dividiendo la frecuencia absoluta de la clase entre el tamaño de la muestra.
f3_s	Frecuencia absoluta teórica	Es necesaria para el cálculo del estadístico χ^2 que se usa para comprobar la bondad del método de generación.

X	Den	n	f1_s	f3_s
0	0,001008	1	0,0011	0,9
1	0,006954	5	0,0054	6,5
2	0,023990	24	0,0258	22,3
3	0,055178	54	0,0581	51,3
4	0,095182	96	0,1032	88,5
5	0,131351	127	0,1366	122,2
6	0,151053	125	0,1344	140,5
7	0,148895	140	0,1505	138,5
8	0,128422	117	0,1258	119,4
9	0,098457	98	0,1054	91,6
10	0,067935	53	0,0570	63,2
11	0,042614	42	0,0452	39,6
12	0,024503	26	0,0280	22,8
13	0,013005	12	0,0129	12,1
14	0,006410	8	0,0086	6,0
15	0,002949	2	0,0022	2,7
16	0,001272	0	0,0000	1,2
17	0,000516	0	0,0000	0,5
	#N/A	0	0,0000	#N/A
	#N/A	0	0,0000	#N/A
	#N/A	0	0,0000	#N/A
	#N/A	0	0,0000	#N/A
	#N/A	0	0,0000	#N/A
	#N/A	0	0,0000	#N/A

Todas las hojas tienen definidos una serie de nombres que permiten, adaptando en cada momento el rango horizontal del gráfico al dominio más probable (99,9%) de la variable aleatoria según la configuración de parámetros elegida por el usuario realizar, el gráfico a partir de la tabla de datos.

Estos nombres³ y su significado son los siguientes:

U	=ALEATORIO()
X	=DESREF(BinNeg!\$B\$3;1; 1 ;1+BinNeg!\$L\$5-BinNeg!\$L\$4;1)
Den	=DESREF(BinNeg!\$B\$3;1; 2 ;1+BinNeg!\$L\$5-BinNeg!\$L\$4;1)
n	=DESREF(BinNeg!\$B\$3;1; 3 ;1+BinNeg!\$L\$5-BinNeg!\$L\$4;1)
f1_s	=DESREF(BinNeg!\$B\$3;1; 4 ;1+BinNeg!\$L\$5-BinNeg!\$L\$4;1)
f3_s	=DESREF(BinNeg!\$B\$3;1; 5 ;1+BinNeg!\$L\$5-BinNeg!\$L\$4;1)

³ Nótese que los rangos aparecen precedidos siempre del nombre de la hoja, en este caso BinNeg!

Anexo 2 Procedimiento genérico para la obtención de v.a. discretas.

Con frecuencia será necesario generar muestras aleatorias que se distribuyan con arreglo a una distribución discreta que ha sido definida por el usuario. Es decir, nos enfrentaremos al problema de simular una distribución especificada en la forma siguiente $\{\mathbf{X}, \mathbf{P}\}$, en donde $\mathbf{X} = \{X_1, X_2, \dots, X_i, \dots, X_n\}$ se refiere al posible conjunto de posibles valores de la variable aleatoria, conjunto cuyos elementos pueden ser números naturales (consecutivos o no) o tal vez códigos, dependiendo de la naturaleza de la variable y de la escala en que ésta ha sido medida, y en donde $\mathbf{P} = \{P_1, P_2, \dots, P_i, \dots, P_n\}$ se refiere a las probabilidades de que X tome precisamente esos valores:

$$\forall i \in \{1, 2, \dots, i, \dots, n\} \quad 0 \leq P_i \leq 1$$

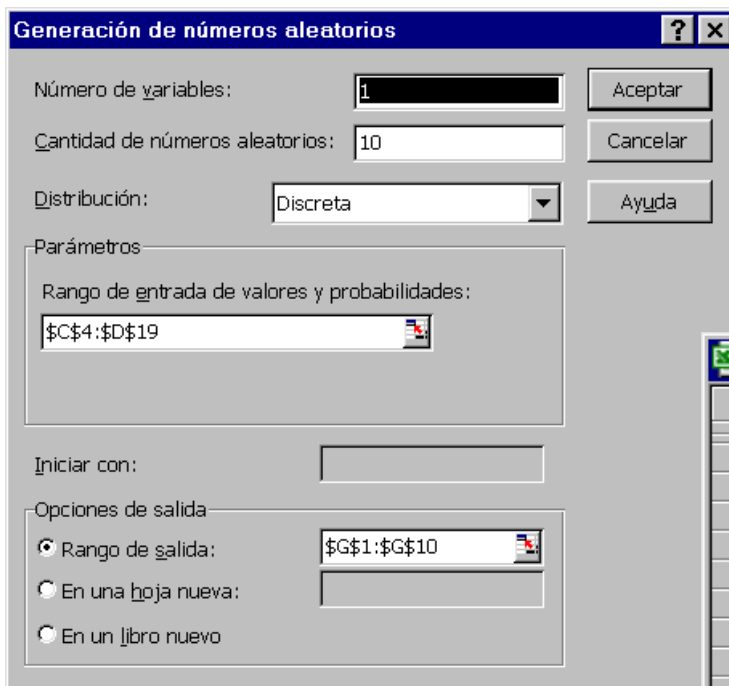
$$\sum_i P_i = 1$$

Excel cuenta con un mecanismo para la generación de variables aleatorias cuya distribución se especifica de esta forma, supongamos que queremos generar muestras de la variable X definida de la forma siguiente:

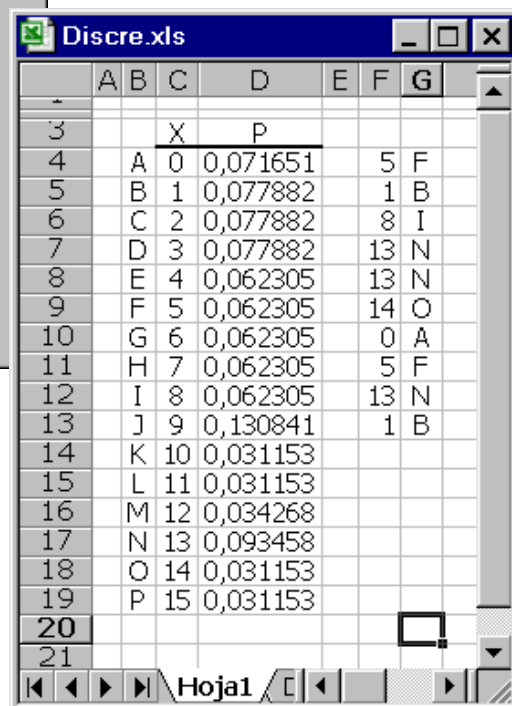
$X = \{A, B, C, \dots, O, P\}$ con probabilidades $\{0.0071651, \dots, 0.031153\}$ (véase la tabla)

Lo primero que debemos tener en cuenta es que Excel no admite valores no numéricos en el dominio de X de manera que será necesario sustituirlos por los enteros naturales, hecho esto invocamos el módulo⁴ de Generación de v.a. en la forma: Herramientas + Análisis de Datos + Generación de números aleatorios". Hecho esto obtendremos un formulario en el que deberemos elegir la opción adecuada, en este caso la opción **Discreta**.

X	P
A	0,071651
B	0,077882
C	0,077882
D	0,077882
E	0,062305
F	0,062305
G	0,062305
H	0,062305
I	0,062305
J	0,130841
K	0,031153
L	0,031153
M	0,034268
N	0,093458
O	0,031153
P	0,031153



Esta opción está "caracterizada por un valor y el rango de probabilidades asociado. El rango debe contener dos columnas. La columna izquierda deberá contener valores y la derecha probabilidades asociadas con el valor de esa fila. La suma de las probabilidades deberá ser 1".



	A	B	C	D	E	F	G
3			X	P			
4	A	0	0,071651		5	F	
5	B	1	0,077882		1	B	
6	C	2	0,077882		8	I	
7	D	3	0,077882		13	N	
8	E	4	0,062305		13	N	
9	F	5	0,062305		14	O	
10	G	6	0,062305		0	A	
11	H	7	0,062305		5	F	
12	I	8	0,062305		13	N	
13	J	9	0,130841		1	B	
14	K	10	0,031153				
15	L	11	0,031153				
16	M	12	0,034268				
17	N	13	0,093458				
18	O	14	0,031153				
19	P	15	0,031153				
20							
21							

La "complimentación" del formulario podría ser como la descrita en esta página. Una vez obtenidos los valores, bastaría con utilizar las funciones de Excel para recuperar los valores reales de la variable a partir de los códigos numéricos obtenidos.

⁴ Para su uso es necesario haber habilitado previamente el módulo de "Análisis de Datos"

Este método tiene la ventaja de ser inmediato y requerir relativamente poco esfuerzo al usuario, sin embargo presenta desventajas, la fundamental derivada del hecho de que la obtención de la muestra no está integrada en la hoja de trabajo. No menos importante que esta es el hecho de que sólo hay posibilidad de generar un pequeño número de variables aleatorias.

Una alternativa a este método se basa en el empleo de las fórmulas matriciales, mecanismo de extraordinaria potencia, que faculta al usuario (avanzado) de Excel para realizar potentes cálculos que no han de plasmarse forzosamente en celda alguna sino que permanecen en memoria hasta que necesitamos de sus resultados.

Antes de ello es conveniente recordar las prestaciones de una de las funciones de librería de Excel menos conocida, nos referimos a la función **PROBABILIDAD**, función que usada convenientemente nos proporciona la distribución de frecuencias acumuladas de una variable a partir de su distribución de frecuencias relativas.

Supongamos el ejemplo anterior con una distribución de probabilidad definida sobre la variable **X** de la forma $\{P_1, P_2, \dots, P_i, \dots, P_n\}$ tal como aparece en la tabla, asignemos en primer lugar el nombre **P** al vector (columna) de las probabilidades:

X	P
A	0,071651
B	0,077882
C	0,077882
D	0,077882
E	0,062305
F	0,062305
G	0,062305
H	0,062305
I	0,062305
J	0,130841
K	0,031153
L	0,031153
M	0,034268
N	0,093458
O	0,031153
P	0,031153

Por comodidad asignemos también el nombre **U** a la fórmula simple **=ALEATORIO()**, y notemos que al asignar al nombre **Sec** la fórmula **=FILA(INDIRECTO("A1:A"& FILAS(P)))** acabamos de crear un vector (columna) que contiene los **n** primeros números naturales siendo **n** la longitud de nuestro vector de probabilidades.

	A	B	C	D	E	F	G	H	I	J	K
3			X	P							
4	A	0	0,071651								
5	B	1	0,077882								
6	C	2	0,077882								
7	D	3	0,077882								
8	E	4	0,062305								
9	F	5	0,062305								
10	G	6	0,062305								
11	H	7	0,062305								
12	I	8	0,062305								
13	J	9	0,130841								
14	K	10	0,031153								
15	L	11	0,031153								
16	M	12	0,034268								
17	N	13	0,093458								
18	O	14	0,031153								
19	P	15	0,031153								

Pues bien, la fórmula **=1-PROBABILIDAD(Sec;P;Xi;FILAS(P))** nos proporciona, la probabilidad de que se produzca un suceso $x < X_i$, por lo que al asignar al nombre **Gen** la **fórmula matricial**:

$$=\{\text{COINCIDIR (U;1-PROBABILIDAD(Sec;P;Xi;FILAS(P));1)}\}$$

obtendremos un valor del vector $X = \{X_1, X_2, \dots, X_i, \dots, X_n\}$ distribuido según $P = \{P_1, P_2, \dots, P_i, \dots, P_n\}$, y si la verdadera naturaleza de la variable (que recordamos hemos codificado de forma numérica) se encuentra en el vector **Nom**, la fórmula

$$=\{\text{INDICE(Nom;+COINCIDIR (U;1-PROBABILIDAD(Sec;P;Xi;FILAS(P));1)1)}\}$$

proporcionará, cada vez que sea invocada en una celda, un valor de **Nom** ↔ **X** distribuido según **P**. El libro [UniDisc.xls](#) contiene una hoja en la que aparecen estos cálculos, su aspecto es el siguiente:

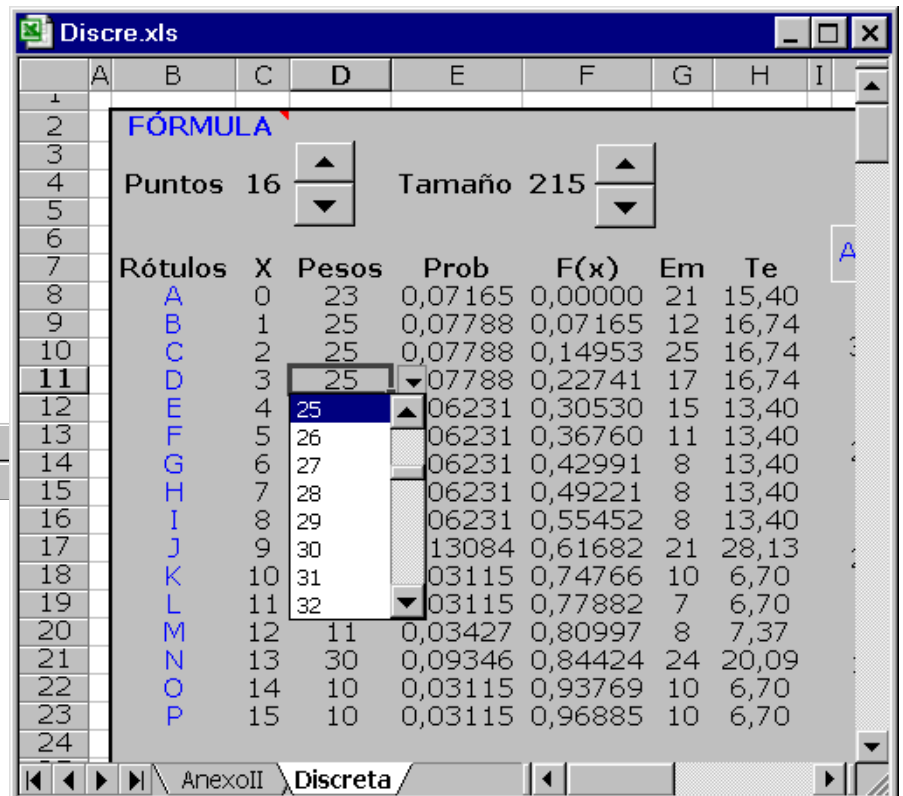
Nom	X	P	Sec	PRO	Xgen	Ngen
A	0	0,071651	1	0,00	13	M
B	1	0,077882	2	0,07	4	D
C	2	0,077882	3	0,15	6	F
D	3	0,077882	4	0,23	10	J
E	4	0,062305	5	0,31	4	D
F	5	0,062305	6	0,37	7	G
G	6	0,062305	7	0,43	2	B
H	7	0,062305	8	0,49	9	I
I	8	0,062305	9	0,55	8	H
J	9	0,130841	10	0,62	7	G
K	10	0,031153	11	0,75	2	B
L	11	0,031153	12	0,78	6	F
M	12	0,034268	13	0,81	1	A
N	13	0,093458	14	0,84	5	E
O	14	0,031153	15	0,94	3	C
P	15	0,031153	16	0,97	10	J
					3	C
					7	G
					16	P
					4	D
					3	C
					6	F
					10	J
					4	D

PRO = 1-PROBABILIDAD(Sec;P;E#;FILAS(P))
Xgen = COINCIDIR(U;1-PROBABILIDAD(Sec;P;Sec;FILAS(P));1)
Ngen = INDICE(Nom;Xgen)

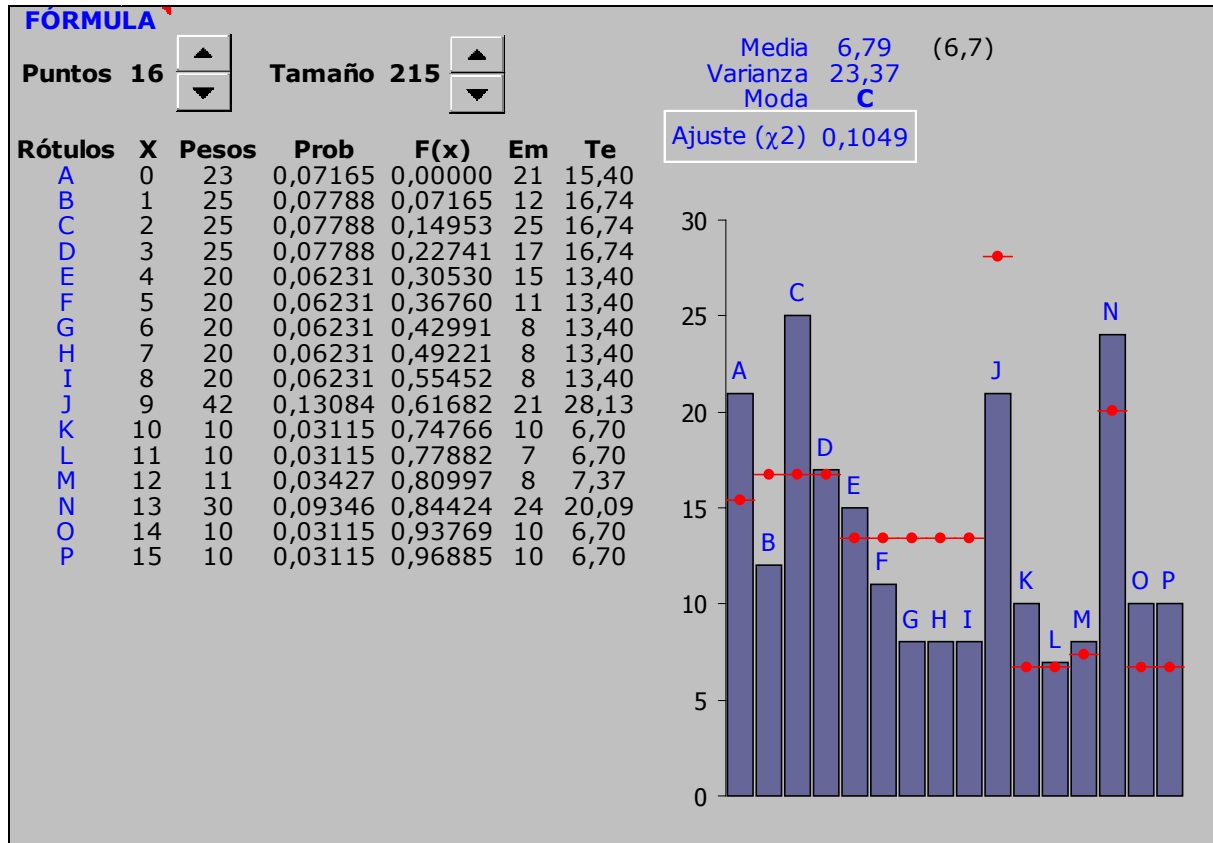
También contiene una hoja que permite la generación de una variable aleatoria especificada por el usuario. En la parte superior aparecen los controles que permiten indicar el número de puntos que tendrá la variable y el tamaño de la muestra que se desea generar.

Puntos 16 **Tamaño 215**

Una vez elegidos los puntos de la variable el usuario puede introducir los valores de P, bien directamente, bien a través de pesos asociados a los diferentes puntos.



El resultado final incluye: la distribución de frecuencias teóricas inducida por los pesos elegidos por el usuario , **Prob F(x)**, las frecuencias absolutas empíricas (**Em**) y teórica (**Te**), el resultado de la prueba de bondad del ajuste y un gráfico de la distribución de la muestra generada, los propios valores que componen la muestra generada tanto en términos de los valores numéricos como de los rótulos a éstos asociados.



Muestra	
12	M
5	F
9	J
14	O
10	K
3	D
5	F
6	G
3	D
6	G
5	F
13	N
13	N
7	H
9	J
9	J
13	N
1	B
4	E
5	F
5	F
1	B

El procedimiento genérico expuesto en este apéndice es el aplicado para la obtención de muestras aleatorias de una variable hipergeométrica.

Para ello hemos sustituido el vector **P**, anteriormente definido por el usuario, por la distribución de probabilidad calculada por Excel para los valores de los parámetros introducidos por el usuario. La fórmula es algo más complicada pero su justificación es análoga a la descrita anteriormente:

```
{=COINCIDIR(ALEATORIO());  
1-PROBABILIDAD(FILA(INDIRECTO("A1:A"&n+1))-1;  
DISTR.HIPERGEOMFILA(INDIRECTO("A1:A"&n+1))-1;n;N;M)  
FILA(INDIRECTO("A1:A"&n+1))-1;n+1);1)  
-1)}
```

Anexo 3 Procedimiento genérico de estimación de parámetros.

El procedimiento estándar para la caracterización de cualquier variable, es decir para la estimación de los parámetros de la distribución a partir de una muestra aleatoria de ésta, se hace habitualmente por el método de máxima verosimilitud.

La literatura está repleta de referencias a este método, su aplicación es muchas veces inmediata. La estimación del parámetro λ de una distribución de Poisson es, simplemente:

$$\hat{\lambda} = \bar{\mathbf{x}}_{(n)}$$

Sin embargo, la caracterización no es siempre una tarea tan sencilla. Consideremos por ejemplo el problema de la estimación máximo verosímil de los parámetros \mathbf{n} y \mathbf{p} de una distribución binomial de la que sólo contamos con una muestra aleatoria de tamaño \mathbf{N} .

La teoría [6] nos dice que las estimaciones de ambos parámetros son aquellos valores que maximizan la función:

$$g_{(n,p)} = \left[\sum_{k=1}^M f_k \text{Ln}(n - k + 1) \right] + [\mathbf{N} \mathbf{n} \text{Ln}(1 - \mathbf{p})] + \left[\mathbf{N} \bar{\mathbf{X}}_{(n)} \text{Ln} \left(\frac{1}{1 - \mathbf{p}} \right) \right]$$

siendo $\mathbf{M} = \max_{\{1 \leq i \leq \mathbf{N}\}} \mathbf{X}_i$ y f_k ($k = 0, 1, \dots, M$) el número de observaciones de la muestra que son menores o iguales que k y sometido a la restricción de que $t \in \{M, M+1, M+2, \dots\}$ lo cual dificulta enormemente el cálculo de $g_{(n,p)}$.

Sin embargo un procedimiento para aproximar los valores de \mathbf{n} y \mathbf{p} puede consistir, simplemente, en elegir aquellos valores que minimicen (o maximicen) un determinado criterio. Por ejemplo, elegir los valores \mathbf{n}, \mathbf{p} que hagan mínima la discrepancia entre las frecuencias absolutas de la muestra y las que cabría esperar bajo la combinación de parámetros elegida, es decir:

$$\hat{\mathbf{n}}, \hat{\mathbf{p}} = \min_{\substack{0 < \mathbf{p} < 1 \\ \mathbf{n} \in \{\mathbf{M}, \mathbf{M}+1, \dots\}}} \mathbf{D}_i^{\mathbf{n}, \mathbf{p}} = \left| \mathbf{N}_i^E - \mathbf{N}_i^{\mathbf{n}, \mathbf{p}} \right|$$

con :

\mathbf{N}_i^E = Las frecuencias absolutas para $i \in \{0, 1, 2, \dots, M\}$ en la muestra,

$\mathbf{N}_i^{\mathbf{n}, \mathbf{p}}$ = Las frecuencias absolutas deducidas de la binomial de parámetros \mathbf{n}, \mathbf{p} es decir:

$$\mathbf{N}_i^{\mathbf{n}, \mathbf{p}} = \sum_{j=0}^{j=i} \binom{\mathbf{n}}{j} \mathbf{p}^j (1 - \mathbf{p})^{\mathbf{n} - j}$$

En principio parecería lógico aprovechar las ventajas de Excel y elegir un criterio del tipo:

$$\hat{\mathbf{n}}, \hat{\mathbf{p}} = \min_{\substack{0 < \mathbf{p} < 1 \\ \mathbf{n} \in \{\mathbf{M}, \mathbf{M}+1, \dots\}}} \mathbf{D}_i^{\mathbf{n}, \mathbf{p}} = \sum_i \frac{(\mathbf{N}_i^E - \mathbf{N}_i^{\mathbf{n}, \mathbf{p}})^2}{\mathbf{N}_i^{\mathbf{n}, \mathbf{p}}}$$

es decir elegir el par \mathbf{n}, \mathbf{p} que minimice la discrepancia en el sentido de la prueba χ^2 , sin embargo, el hecho de que Excel devuelva #NA's cuando algún $\mathbf{N}_i^{\mathbf{n}, \mathbf{p}}$ es nulo hace complicada la ejecución de este criterio.

El fichero [Binomial.xls](#) contiene una segunda hoja en la que se ha implementado el procedimiento de estimación anterior. La hoja contiene una tabla en la en las filas se ha discretizado el dominio de \mathbf{p} en intervalos de longitud 0,025 y las columnas corresponden a los primeros valores de la secuencia $\{M, M+1, M+2, \dots\}$ siendo $\mathbf{M} = \max_{\{1 \leq i \leq \mathbf{N}\}} \mathbf{X}_i$.

En cada celda, correspondiente al valor de la discrepancia para el valor candidato de \mathbf{p} (filas) y el valor candidato de \mathbf{n} (columnas), se calcula la discrepancia:

$$\mathbf{D}_i^{\mathbf{n}, \mathbf{p}} = \left| \mathbf{N}_i^E - \mathbf{N}_i^{\mathbf{n}, \mathbf{p}} \right|$$

utilizando la fórmula matricial siguiente:

$$\text{MAX}(\text{ABS}(\text{DISTR.BINOM}(\text{Sec};n;p;0)-\text{FRECUENCIA}(\text{D};M)/\text{CONTAR}(\text{D})))$$

habiéndose definido previamente los nombres:

- D = INDIRECTO(Datos a ajustar)
- M = FILA(INDIRECTO("A1:A"&MAX(D)+1))-1
- Sec = FILA(INDIRECTO("A1:A"&FILAS(M)+1))-1

	16	17	18	19	20	21	22	23	24	25	26
0,101	16	17	18	19	20	21	22	23	24	25	26
0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
0,025	0,667	0,650	0,634	0,618	0,603	0,588	0,573	0,559	0,545	0,531	0,518
0,050	0,440	0,418	0,397	0,377	0,358	0,341	0,324	0,307	0,292	0,277	0,264
0,075	0,287	0,266	0,246	0,227	0,210	0,195	0,180	0,166	0,154	0,142	0,132
0,100	0,185	0,167	0,150	0,135	0,122	0,109	0,098	0,089	0,087	0,095	0,102
0,125	0,118	0,103	0,090	0,088	0,097	0,106	0,114	0,120	0,126	0,131	0,136
0,150	0,092	0,104	0,113	0,121	0,128	0,134	0,139	0,143	0,146	0,149	0,152
0,175	0,121	0,129	0,135	0,141	0,145	0,149	0,152	0,155	0,157	0,159	0,160
0,200	0,139	0,144	0,149	0,152	0,155	0,157	0,159	0,161	0,162	0,163	0,164
0,225	0,150	0,154	0,156	0,159	0,161	0,162	0,163	0,164	0,164	0,165	0,165
0,250	0,157	0,159	0,161	0,162	0,163	0,164	0,165	0,165	0,166	0,166	0,166
0,275	0,161	0,162	0,164	0,164	0,165	0,165	0,166	0,166	0,166	0,166	0,166
0,300	0,163	0,164	0,165	0,166	0,166	0,166	0,166	0,166	0,166	0,167	0,167
0,325	0,165	0,165	0,166	0,166	0,166	0,166	0,166	0,167	0,167	0,167	0,167
0,350	0,166	0,166	0,166	0,166	0,166	0,166	0,167	0,167	0,167	0,167	0,167
0,375	0,166	0,166	0,166	0,167	0,167	0,167	0,167	0,167	0,167	0,167	0,167

Mediante formato condicional se indican los pares de valores n, p que son menores al 110% de la mínima discrepancia encontrada, en este caso (la figura muestra sólo una parte de la hoja) los valores candidatos serían $\{(n=16;p=0,15) ; (18 \leq n \leq 20;p=0,125) ; (22 \leq n \leq 25;p=0,100) ; \dots\}$.

	16	17	18	19	20	21	22	23	24
0,101	16	17	18	19	20	21	22	23	24
0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
0,025	0,667	0,650	0,634	0,618	0,603	0,588	0,573	0,559	0,545
0,050	0,440	0,418	0,397	0,377	0,358	0,341	0,324	0,307	0,292
0,075	0,287	0,266	0,246	0,227	0,210	0,195	0,180	0,166	0,154
0,100	0,185	0,167	0,150	0,135	0,122	0,109	0,098	0,089	0,087
0,125	0,118	0,103	0,090	0,088	0,097	0,106	0,114	0,120	0,126
0,150	0,092	0,104	0,113	0,121	0,128	0,134	0,139	0,143	0,146
0,175	0,121	0,129	0,135	0,141	0,145	0,149	0,152	0,155	0,157
0,200	0,139	0,144	0,149	0,152	0,155	0,157	0,159	0,161	0,162
0,225	0,150	0,154	0,156	0,159	0,161	0,162	0,163	0,164	0,164
0,250	0,157	0,159	0,161	0,162	0,163	0,164	0,165	0,165	0,166
0,275	0,161	0,162	0,164	0,164	0,165	0,165	0,166	0,166	0,166
0,300	0,163	0,164	0,165	0,166	0,166	0,166	0,166	0,166	0,166
0,325	0,165	0,165	0,166	0,166	0,166	0,166	0,166	0,167	0,167
0,350	0,166	0,166	0,166	0,166	0,166	0,167	0,167	0,167	0,167
0,375	0,166	0,166	0,166	0,167	0,167	0,167	0,167	0,167	0,167

BIBLIOGRAFÍA

- [1]. Bennet B.S. **"Simulation Fundamentals"**, Prentice Hall (1995).
- [2]. Bratley, P., Fox, B. y Schrage, L. **"A Guide to Simulation"**, Springer. (1987)
- [3]. Fishman, G.S. **"Monte Carlo: Concepts, Algorithms and Applications"**, Springer (3rd ed.) (1999).
- [4]. Gordon G. **"System Simulation"** Editorial Prentice-Hall (1978).
- [5]. Kleijnen, J. y Van Groenendaal, W. **"Simulation: A Statistical Perspective"**, Wiley(1992).
- [6]. Law A. M., W. D. Kelton **"Simulation Modeling & Analysis"** Ed. McGraw-Hill (1984).
- [7]. MacDougall M.H. **"Simulating Computer Systems Techniques and Tools"** MIT Press (1987).
- [8]. Motwani y Raghavan **"Randomized Algorithms"**, Cambridge U.P (1995).
- [9]. Neelamkavil, F. **"Computer Simulation and Modelling"**, Wiley (1988).
- [10]. P. Bratley, B. L. Fox, L. E. Schrage **"A Guide to Simulation"** Springer-Verlag (1987).
- [11]. Raj Jain **"The Art of Computer Systems Performance Analysis"** Ed. Wiley (1991).
- [12]. Ríos Insua et al. **"Simulación métodos y aplicaciones"** Ed. Ra-Ma, Madrid (1997)
- [13]. Ripley, B. **"Stochastic Simulation"**, Wiley (1987).
- [14]. Ross, S. **"A Course in Simulation"**, MacMillan (1990).
- [15]. Thompson, J.R. **"Simulation: A modeler's Approach"**. John Wiley, 2000.
- [16]. Ziegler, B. **"Theory of Modelling and Simulation"**, Wiley (1976).