Variables Aleatorias y Principios de Simulación

http://humberto-r-alvarez-a.webs.com

Conceptos de probabilidad

- La Teoría de Probabilidad trata fenómenos que pueden ser modelados por experimentos cuyos resultados están gobernados por el azar (se denominan *experimentos aleatorios*). Estos experimentos aleatorios están caracterizados por
  - Los experimentos son repetibles bajo idénticas condiciones
  - El resultado de un experimento es impredecible
  - Si el experimento se realiza un gran número de veces, el resultado exhibe un cierta regularidad estadística (se observa un comportamiento promedio).

Probabilidades

- ¿En qué consisten las probabilidades?
- Indican incertidumbre acerca de un evento que:
  - Ocurrió en el pasado
  - Ocurre en el presente
  - Ocurrirá en el futuro

Enfoques de probabilidad

- Clásico o escuela objetiva
- Frecuencias relativas
- Personalista o subjetivo

Fuentes de las probabilidades

- Historia del pasado
- Juicio subjetivo
- Distribuciones teóricas

La Variable Aleatoria

- Una variable aleatoria es una función que asocia un número real con cada elemento del espacio muestral de un experimento
- **Variable aleatoria discreta:** si se puede contar su conjunto de resultados posible o si están relacionadas con el conjunto de números enteros.
- **Variable aleatoria continua**: Cuando puede tomar valores en la escala continua o está relacionada con un conjunto continuo de valores.

Distribuciones de probabilidad

- Una variable aleatoria toma cada uno de sus valores con cierta probabilidad.
- La función de probabilidad es la representación de todas las probabilidades de una variable aleatoria X mediante una expresión matemática tal que

$$P(X = x) = f(x)$$

Distribuciones de Probabilidad discreta

Distribuciones de probabilidad discretas

- El conjunto de pares ordenados $(x, f(x))$ es una función de probabilidad discreta o distribución de probabilidad discreta de la variable discreta X, si para cada resultado x
  1. $f(x) \geq 0$
  2. $\Sigma_x f(x) = 1$
  3. $P(X = x) = f(x)$
- La distribución acumulada F(x) de una variable aleatoria discreta X con distribución de probabilidad $f(x)$ es:

$$F(x) = P(X \leq x) = \Sigma_{t \leq x} f(t) \quad \text{para } -\infty \leq x \leq \infty$$

Media de una variable aleatoria

- La media de una variable aleatoria es el valor más esperado de dicha variable.

$$\mu = E(X) = \sum_{x} x f(x)$$

Varianza de una variable aleatoria

- Sea X una variable aleatoria con distribución de probabilidad f(x) y media μ, la varianza de X será:

$$\sigma^2 = E\left[(X-\mu)^2\right] = \sum_{x} (x-\mu)^2 f(x)$$

Donde la desviación estándar σ será la raiz de la varianza

Distribución uniforme discreta

- Si la variable aleatoria discreta X toma valores discretos x_1, x_2, ….., x_k con igual probabilidad, entonces estos valores están distribuidos en función a la Distribución Uniforme Discreta:

$$P(X = x) = f(x;k) = \frac{1}{k}$$

$$\mu = \frac{\sum_{i=1}^{k} x_i}{k}$$

$$\sigma^2 = \frac{\sum_{i=1}^{k} (x_i - \mu)^2}{k}$$

El Proceso de Bernoulli

- El experimento consiste en n pruebas iguales que se repiten
- Cada prueba produce un resultado que se puede clasificar como éxito o fracaso
- La probabilidad de un éxito, p, permanece constante en cada prueba
- Cada prueba es independiente

La Distribución Binomial

- En experimento de Bernoulli puede tener como resultado un éxito con probabilidad p y un fracaso con probabilidad 1-p. Entonces la distribución de probabilidad de la variable aleatoria X, el número de éxitos en n pruebas independientes es:

$$P(X = x) = f(x;n,x) = \binom{n}{x} p^x q^{n-x}$$

donde

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

$$\mu = np$$

$$\sigma^2 = npq$$

$$P(x \leq x) = \sum_{k=0}^{x} \binom{n}{k} p^k q^{n-k}$$

Proceso de Poisson

- Son aquellos experimentos donde los resultados ocurren durante un intervalo dado o en una región específica.
- Los resultados que ocurren en un intervalo son independientes de los resultados en otro intervalo o región.
- La probabilidad de que ocurra un solo resultado durante un intervalo muy corto es proporcional a la longitud del intervalo y no depende del número de resultados en otros intervalos
- La probabilidad de que se den resultados simultáneos en un intervalo es despreciable.

Distribución de Poisson

- Sea X la variable aleatoria asociada con la ocurrencia de eventos en un proceso de Poisson. Esta variable estará distribuida de acuerdo a la siguiente función:

$$P(X = x) = \frac{e^{-1/\lambda}(1/\lambda)^{x}}{x!}, \text{para } x = 1, 2, 3, ...$$

$$P(X \leq x) = \sum_{k=0}^{x} \frac{e^{-1/\lambda}(1/\lambda)^{k}}{k!}$$

$$\mu = \sigma \quad = 1/\lambda$$

Donde $1/\lambda$ es el número promedio de eventos por unidad de tiempo

Distribuciones de Probabilidad continuas

Distribuciones de probabilidad continuas

- El conjunto de pares ordenados $(x, f(x))$ es una función de probabilidad continua o función de densidad de probabilidad de la variable aleatoria continua X, definida en el conjunto de números reales $\Re$, si:
  1. $f(x) \geq 0$, para toda $x \in \Re$
  2. $\int_{\Re} f(x) = 1$
  3. $P(a < x < b) = \int_{a < x < b} f(x)dx$
- La distribución acumulada F(x) de una variable aleatoria continua X con distribución de probabilidad $f(x)$ es:

$$F(x) = P(X \leq x) = \int_{t \leq x} f(t)dt \quad \text{para } -\infty \leq x \leq \infty$$

Media de una variable aleatoria

- La media de una variable aleatoria es el valor más esperado de dicha variable.

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

Varianza de una variable aleatoria

- Sea X una variable aleatoria con distribución de probabilidad f(x) y media μ, la varianza de X será:

$$\sigma^2 = E\left[(X-\mu)^2\right] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

Donde la desviación estándar σ será la raiz de la varianza

Distribución uniforme

- Una variable aleatoria continua X tendrá una distribución uniforme en el intervalo [a, b] de la forma:

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

$$P(X \leq x) = \int_0^x \frac{dy}{b-a}$$

$$\mu = \frac{a+b}{2}, \sigma^2 = \frac{(b-a)^2}{12}$$

Distribución exponencial

- La variable aleatoria continua X tiene una **distribución exponencial** con parámetro λ si su función de probabilidad está dada por:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{en cualquier otro caso} \end{cases}$$

$$P(X \leq x) = \int_0^x \lambda e^{-\frac{1}{\lambda} y} dy$$

$$P(0 \leq X \leq x) = 1 - e^{-\frac{1}{\lambda} x}$$

$$\mu = \frac{1}{\lambda}, \quad \sigma = \frac{1}{\lambda}$$

Donde $1/\lambda$ es el número promedio de eventos por unidad de tiempo

Distribución normal

- Es la distribución más importante en el campo de la estadística
- La curva normal describe aproximadamente muchos fenómenos que ocurren en la naturaleza
- Desarrollada en 1733 por Abraham DeMoivre
- Aunque fue Karl Fiedrich Gauss quien demostró su aplicabilidad

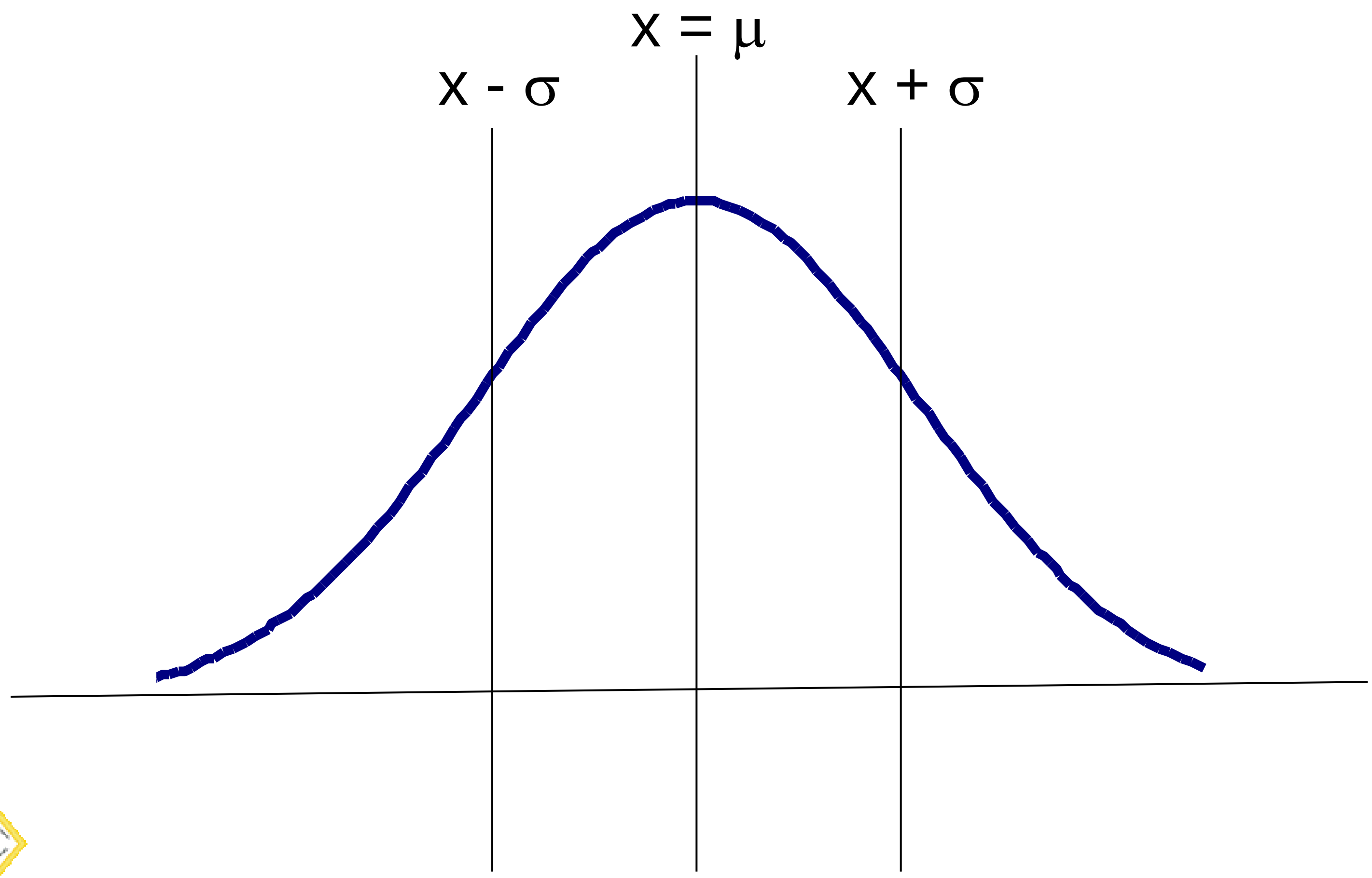
Distribución normal

- La función de probabilidad de la variable aleatoria normal X, con media μ y varianza σ^2 está dada por:

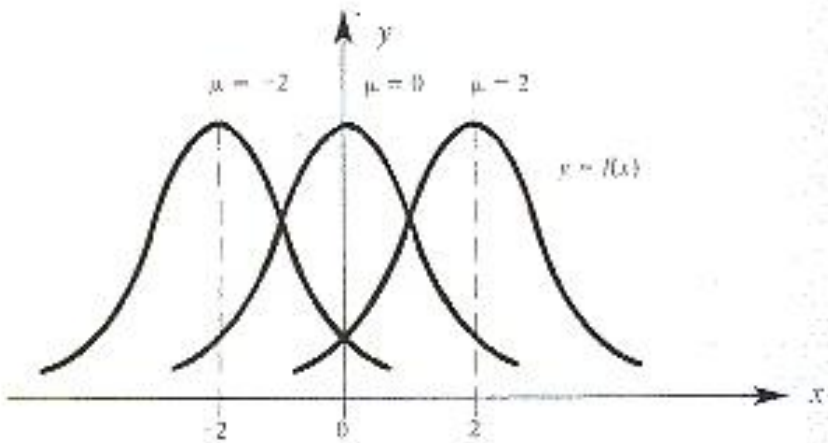
$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}, \quad -\infty < x < \infty$$

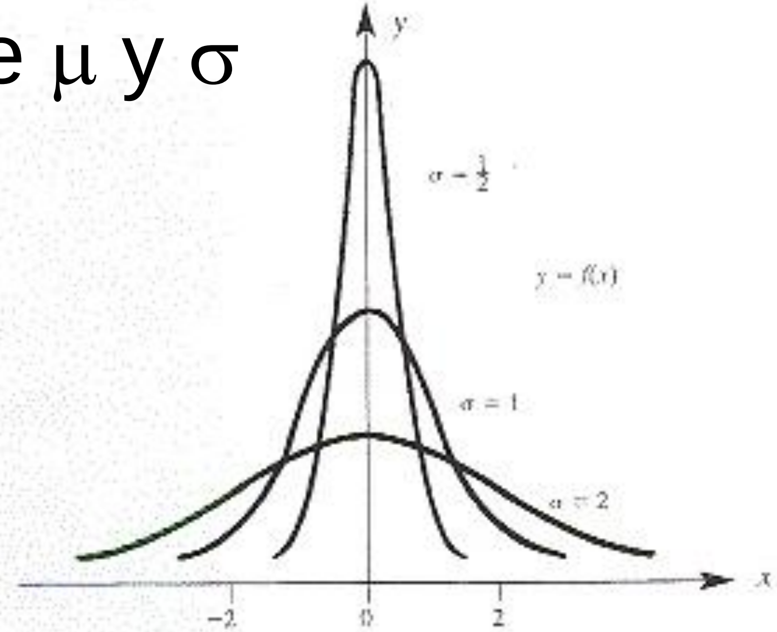
Características de la curva normal

- La variable aleatoria normal X se expresa como $N(\mu, \sigma^2)$
- La moda está localizada en el punto donde $x = \mu$
- La curva es simétrica en el eje de la media μ
- Sus puntos de inflexión están en $x = \mu \pm \sigma$
- Se aproxima al eje horizontal de manera asintótica en ambos lados
- Su área total es igual a 1


X = μ
X - σ
X + σ

Efectos de μ y σ





Áreas bajo la curva normal

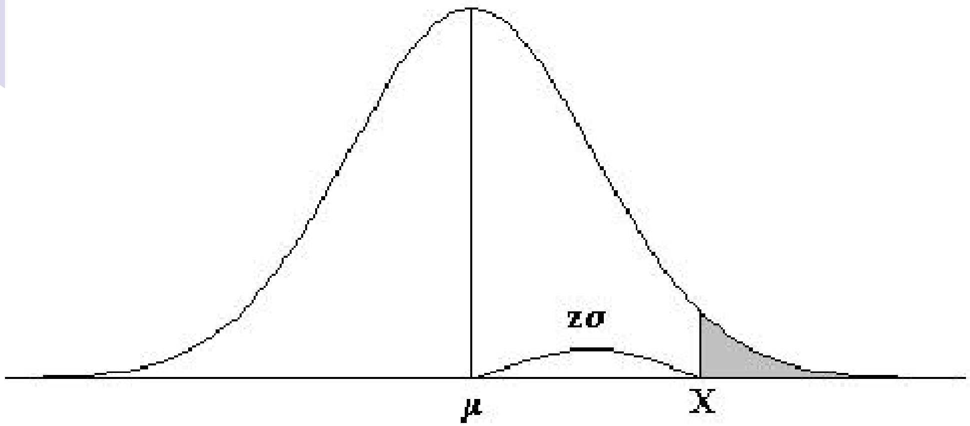
- Es matemáticamente difícil calcular las probabilidades o el área bajo la curva normal:

$$P(x_1 < X < x_2) = \frac{1}{\sqrt{2\pi\sigma}} \int_{x_1}^{x_2} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} dx$$

- Se determina el valor normalizado estándar tal que N(0,1)

$$Z = \frac{x-\mu}{\sigma}$$

Áreas bajo la curva normal



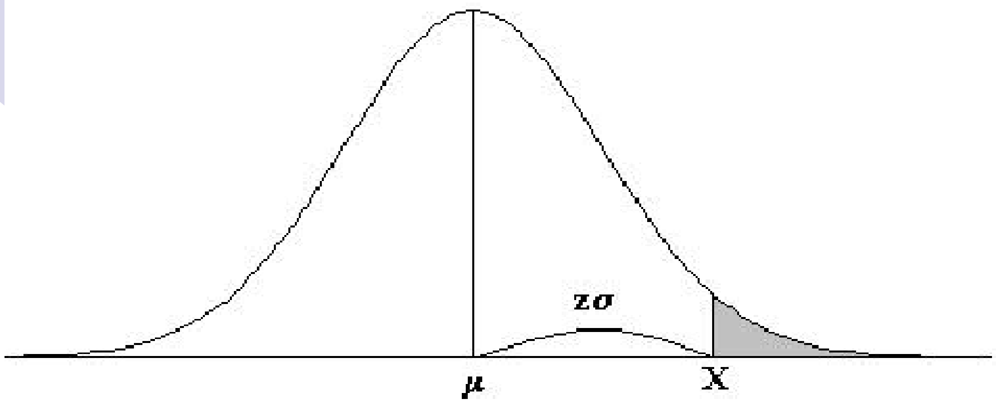
Ejemplo:

$$Z = \frac{X - \mu}{\sigma}$$

P [Z > 1] = 0.1587
P [Z > 1.96] = 0.0250

Desv. normal x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681

Áreas bajo la curva normal



Ejemplo:

$$Z = \frac{X - \mu}{\sigma}$$

$P\,[Z > 1] = 0.1587$

$P\,[Z > 1.96] = 0.0250$

Desv. normal x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010

Simulando procesos estocásticos estacionarios

Funciones generadoras

$$Exponencial:$$

$$x_i = -1/\lambda \ln(R_i)$$

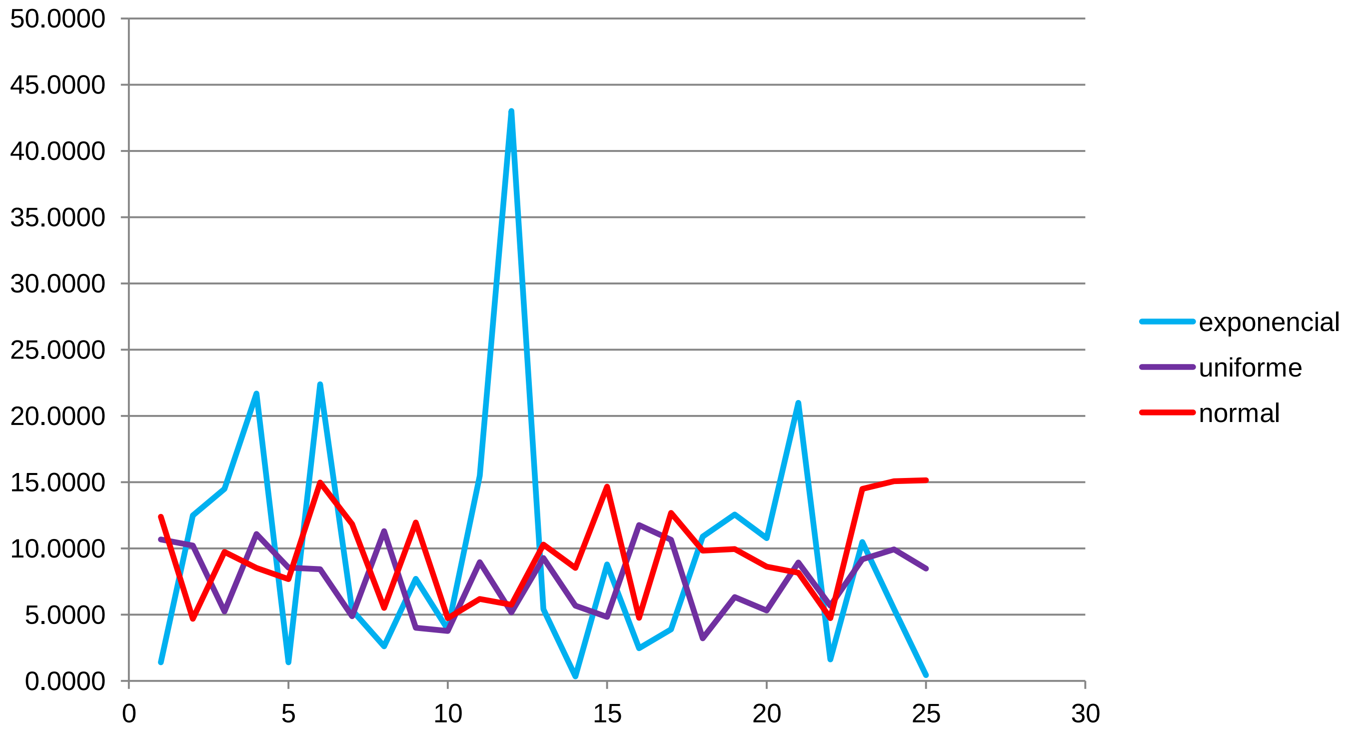
$$Uniforme:$$

$$x_i = a + (b - a) \times R_i$$

$$Normal:$$

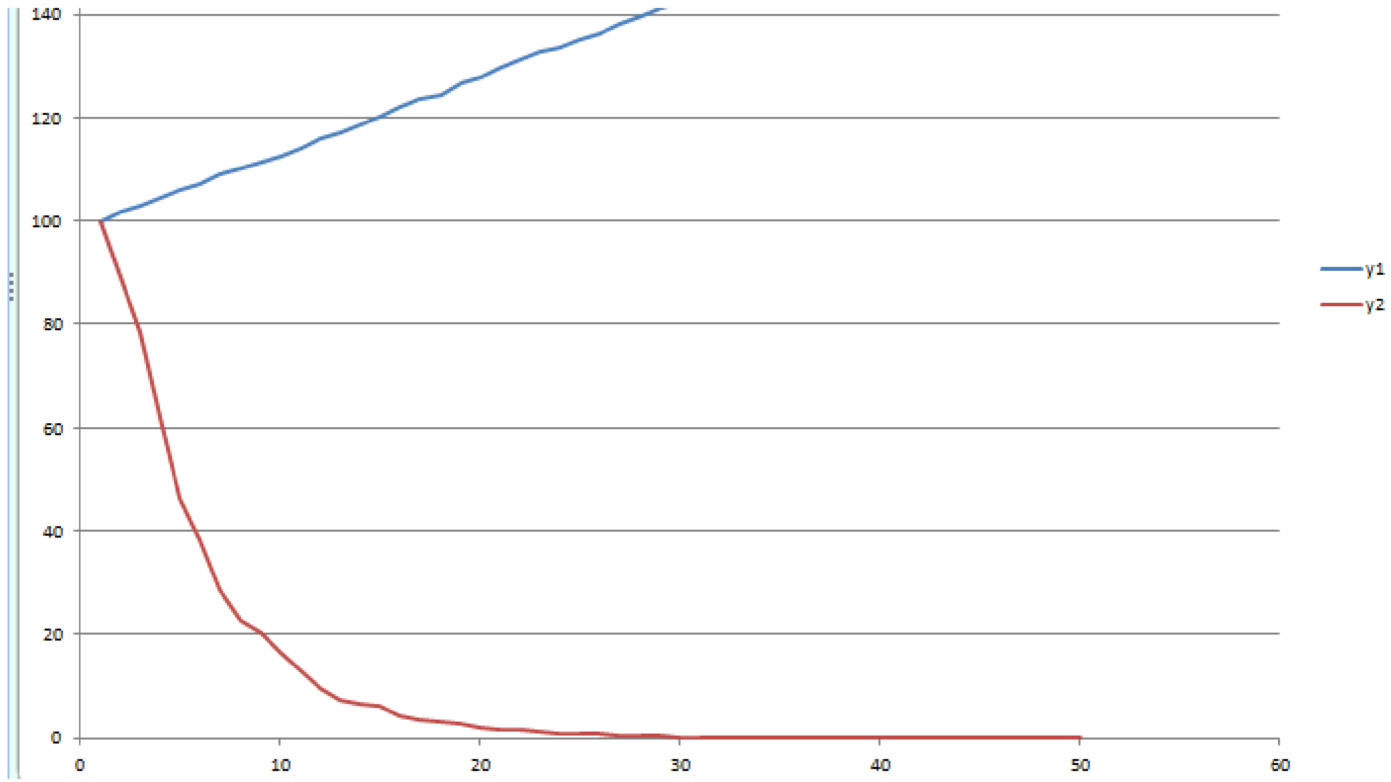
$$x_i = \mu + (6R_i - 3)\sigma$$

	a,b	μ, σ
1/λ	3	10
10	12	2
$x_i = -1/\lambda \ln(R_i)$	$x_i = a + (b-a) \times R_i$	$x_i = \mu + (6R_i - 3)\sigma$
exponencial	uniforme	normal
=-$B$3*LN(ALEATORIO())	=$C$2+($C$3-$C$2)*ALEATORIO()	=$D$2+(6*ALEATORIO()-3)*$D$3

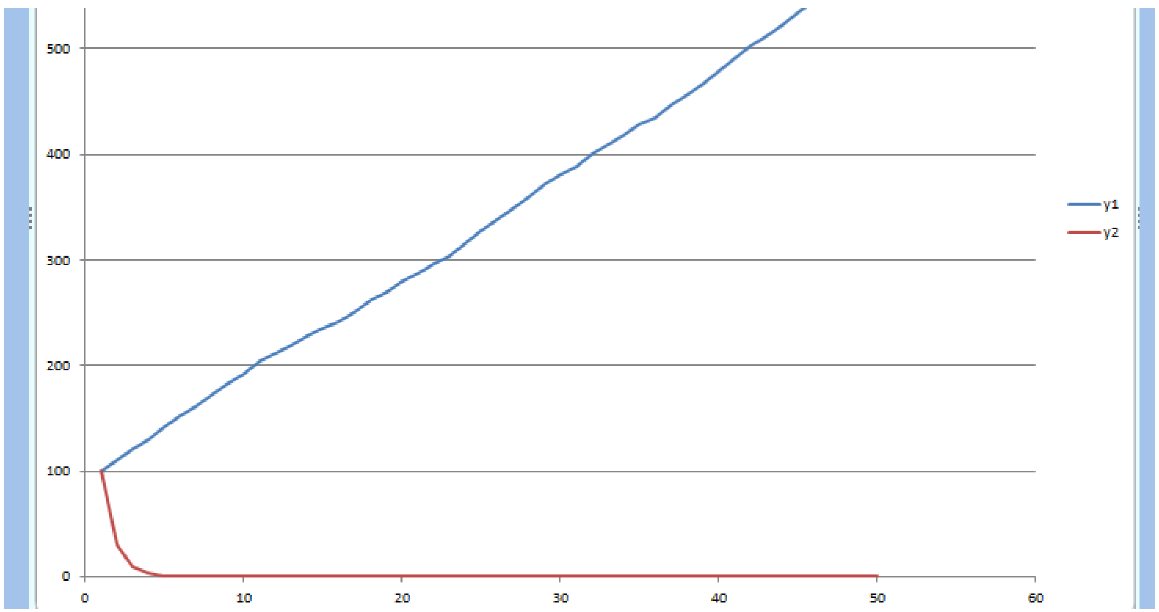


Crecimientos aleatorios

$$y_{i+1} = (1 + U(a,b)) + y_i \qquad y_{i+1} = \frac{y_i}{\sqrt{1 + U(a,b)}}$$



	A	B	C
1			
2		a	5.2
3		b	11.5
4			
5		$y_{i+1} = (1 + U(a,b)) + y_i$	$y_{i+1} = \frac{y_i}{\sqrt{1 + U(a,b)}}$
6	i	y1	y2
7	1	100	100
8	2	=(1+$C$2+($C$3-$C$2)*ALEATORIO())+B7	=C7/RAIZ(1+$C$2+($C$3-$C$2)*ALEATORIO())

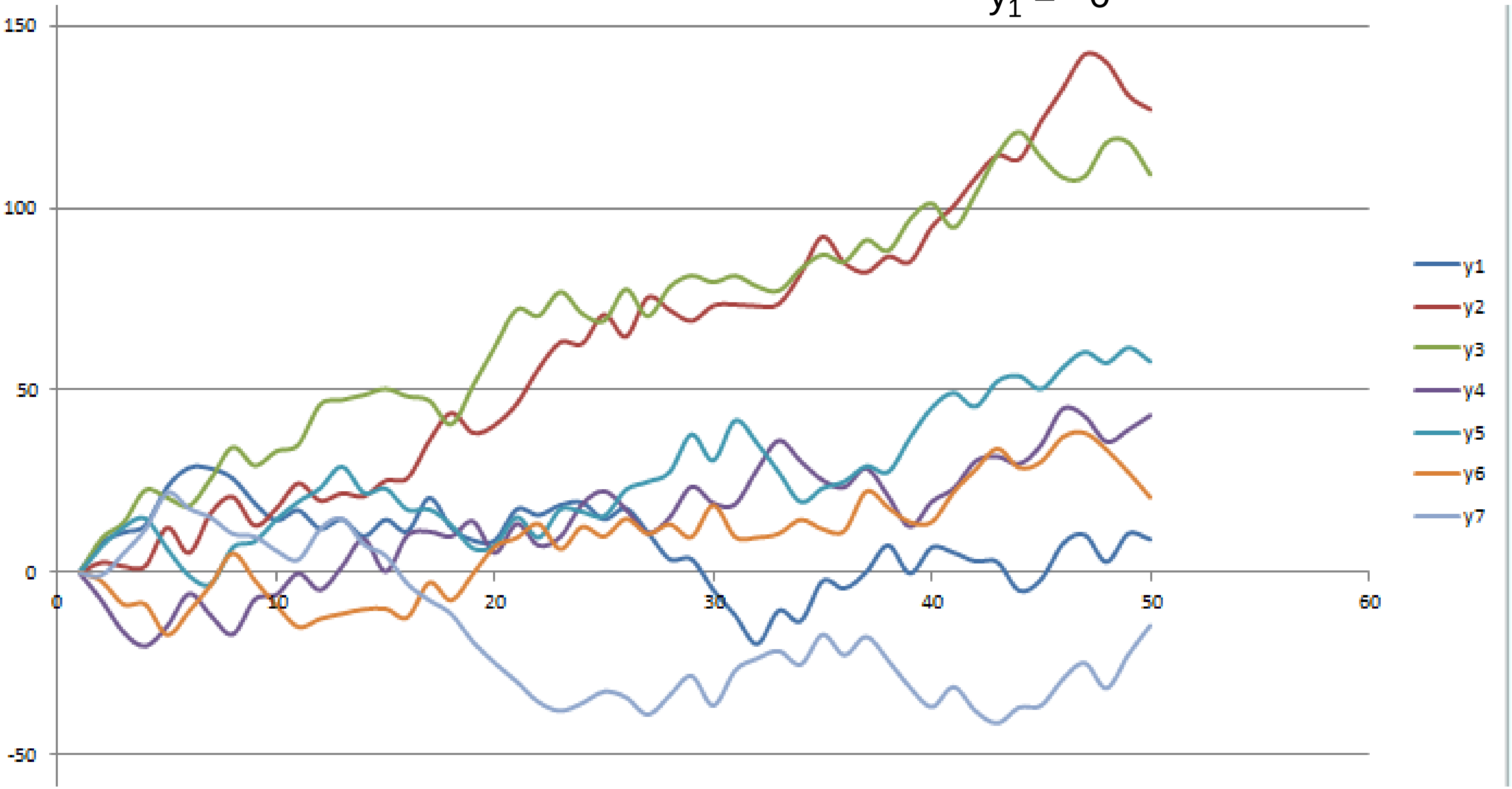


Paseo aleatorio

- Se hacen siete corridas aleatorias para la función

$$y_{i+1} = (1 + U(a,b)) + y_i$$

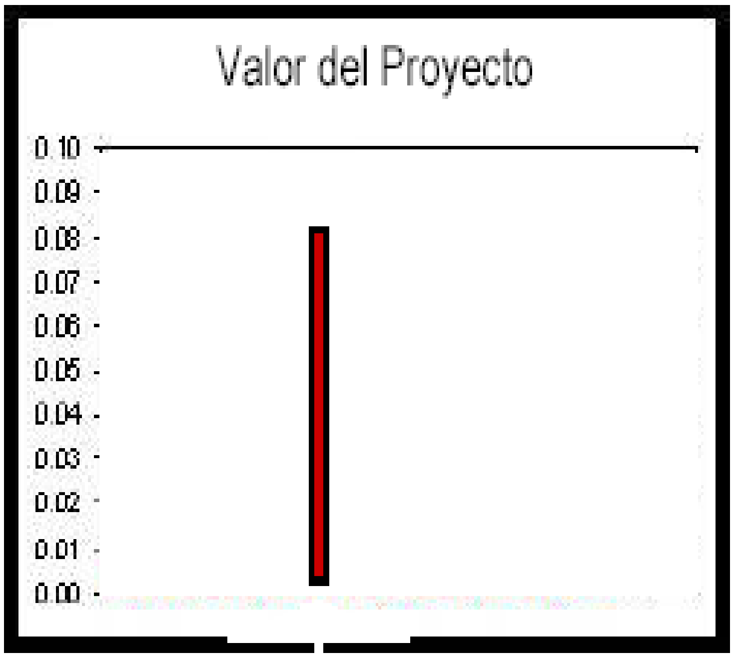
Con parámetros:
$a = -10$
$b = 10$
$y_1 = 0$



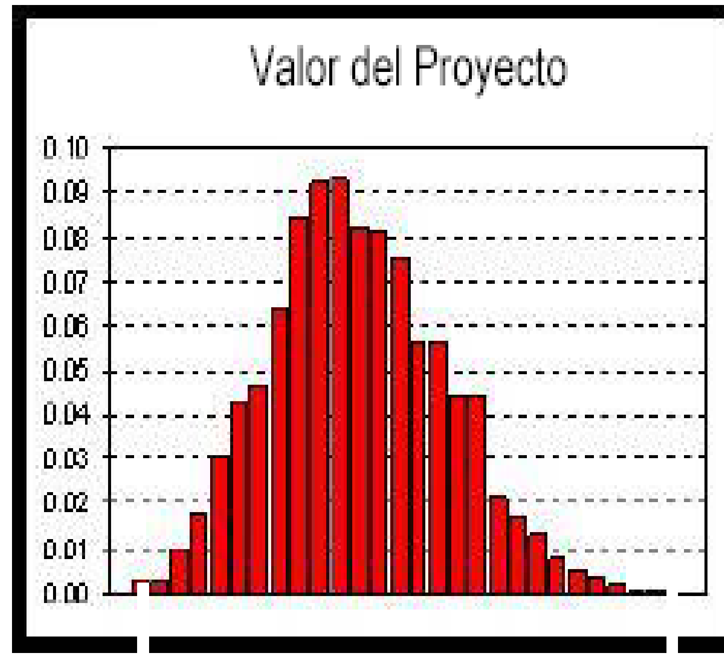
Modelado estocástico

- Cuando se realiza un análisis determinístico a un modelo, una serie de supuestos y variables producen un resultado de valor único.
- Mientras que un análisis estocástico o probabilístico le da al analista un rango de valores como resultado.
- Los resultados estocásticos son mucho más realistas que estimados de valor único ya que se enfocan tanto en la probabilidad de ocurrencia como en las consecuencias o impactos de los riesgos potenciales.

Resultado de Modelo Estático



Resultados de Modelo Probabilístico



Métodos Monte Carlo

- Abarcan una colección de técnicas que permiten obtener soluciones de problemas matemáticos o físicos por medio de pruebas aleatorias repetidas.
- En la práctica, las pruebas aleatorias se sustituyen por resultados de ciertos cálculos realizados con números aleatorios.

Simulación de Monte Carlo

- Es una técnica de análisis de riesgos que incorpora múltiples simulaciones de resultados con la variabilidad de elementos individuales para producir una distribución de resultados potenciales.
- Para cada simulación, la herramienta de simulación Monte Carlo escoge al azar un valor para cada evento de riesgo dentro de su rango de valores posibles, pero de acuerdo con la probabilidad de ocurrencia de cada uno de éstos.
- Luego se combinan los valores escogidos al azar para generar un solo resultado para una simulación. Este proceso se repite un cierto número de veces (típicamente más de 1,000 iteraciones), y se produce un rango de resultados potenciales igualmente probables.

Historia

- El método fue llamado así por el principado de Mónaco por ser “la capital del juego de azar”, al tomar una ruleta como un generador simple de números aleatorios.
- El nombre y el desarrollo sistemático de los métodos de Monte Carlo datan de circa 1944 con el desarrollo de la computadora.
- El uso real de los métodos de Monte Carlo como una herramienta de investigación, proviene del trabajo de la bomba atómica durante la Segunda Guerra Mundial.
- Este trabajo involucraba la simulación directa de problemas probabilísticos de hidrodinámica concernientes a la difusión de neutrones aleatorios en material de fusión.

Algoritmo

- *Determinar la/s V.A. y sus distribuciones acumuladas(F)*
- *Generar un número aleatorio*
- *uniforme $\in (0,1)$.*
- *Determinar el valor de la V.A. para el número aleatorio generado de acuerdo a las clases que tengamos.*

Iterar tantas veces como muestras necesitamos

- *Calcular media, desviación estándar error y realizar el histograma.*
- *Analizar resultados para distintos tamaños de muestra.*

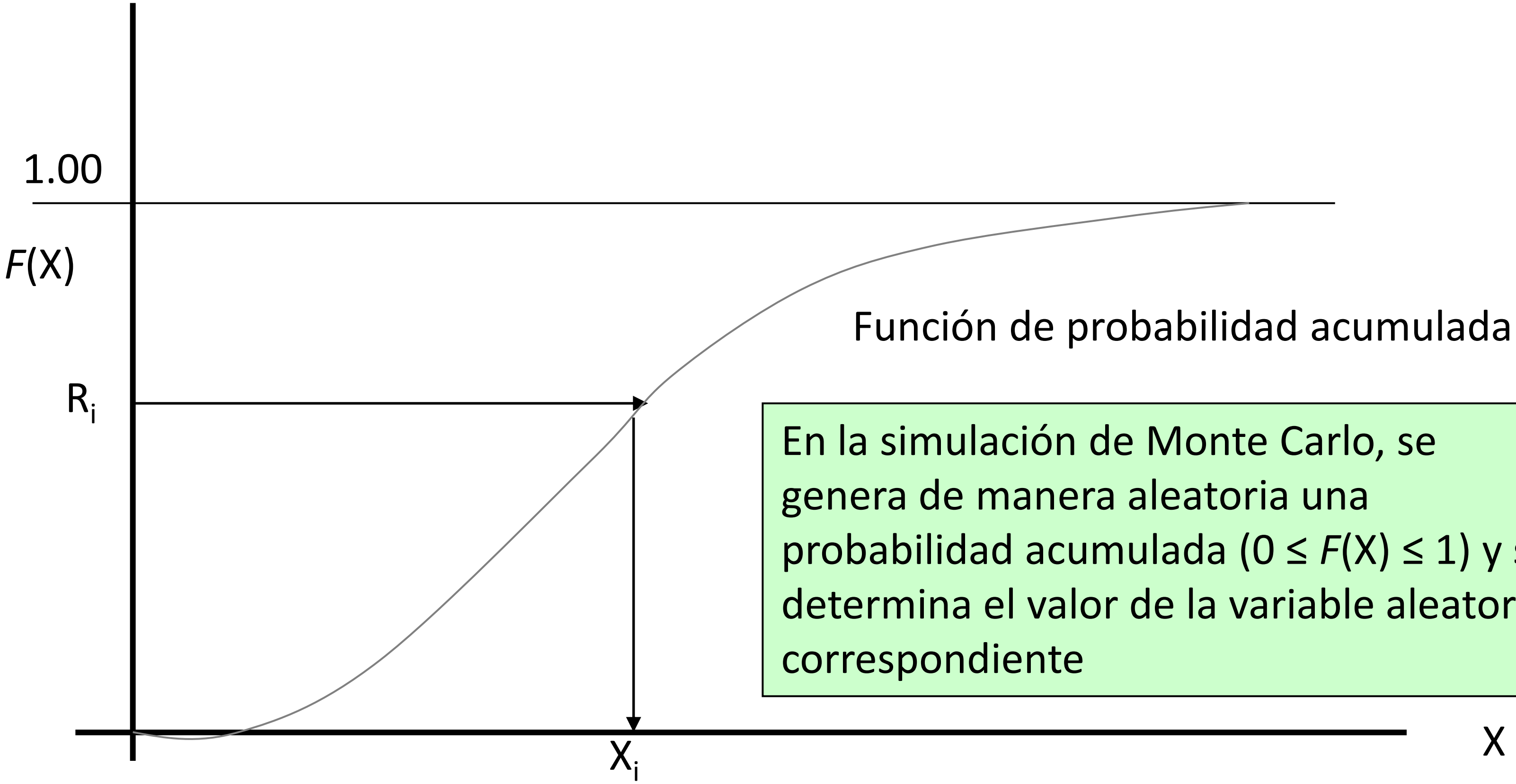
Números aleatorios

- Deben tener igual probabilidad de salir elegidos.
- No debe existir correlación serial
- Se generan por tablas (Rand 1955), o por dispositivos especiales: ruleta.
- En la práctica se utilizan algoritmos y se generan números pseudo aleatorios.
  - Sustituyen a los números aleatorios.
  - Se generan por algoritmos o fórmulas.
  - Se debe asegurar la existencia de secuencias largas y densas.

Generación de números pseudo aleatorios

- *Centros Cuadrados:*
  - $44^2 = 1936 \Rightarrow 93$
- *Métodos Congruenciales:*
  - $x_n = (ax_{n-1} + c)\ (mod\ m$
- *Transformación Inversa*
  - $x = F^{-1}(x)$ siendo $F(x) = Prob(X <= x)$

La función de probabilidad invertida



En la simulación de Monte Carlo, se genera de manera aleatoria una probabilidad acumulada ($0 \leq F(X) \leq 1$) y se determina el valor de la variable aleatoria correspondiente

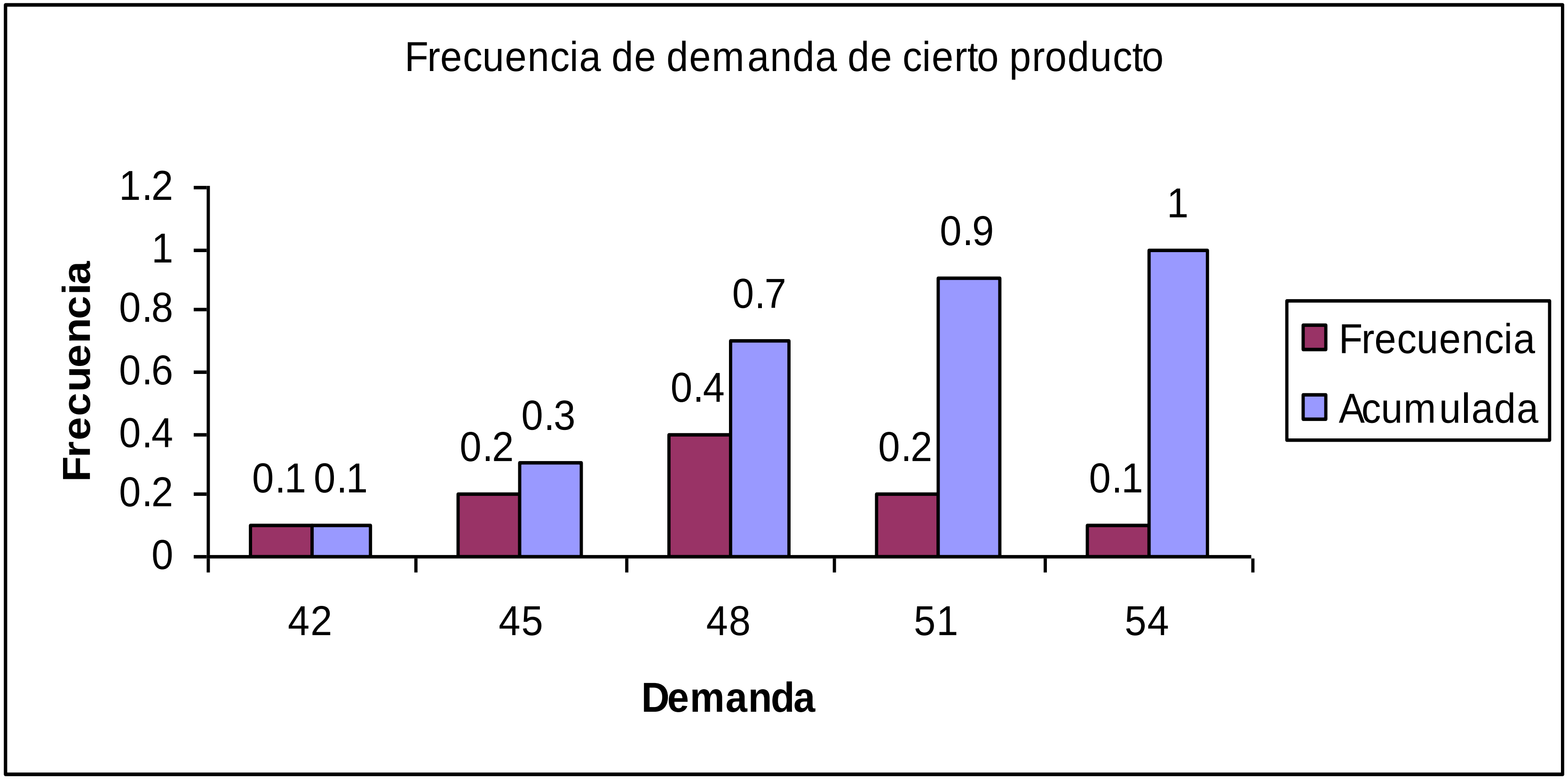
Ejemplo: Se tiene la siguiente distribución de la demanda (días):

Demanda	Frecuencia
42	10
45	20
48	40
51	20
54	10

Modelar estocásticamente su comportamiento y encontrar sus parámetros principales

Frecuencia acumulada

Demanda	Frecuencia	Acumulada
42	0.1	0.1
45	0.2	0.3
48	0.4	0.7
51	0.2	0.9
54	0.1	1

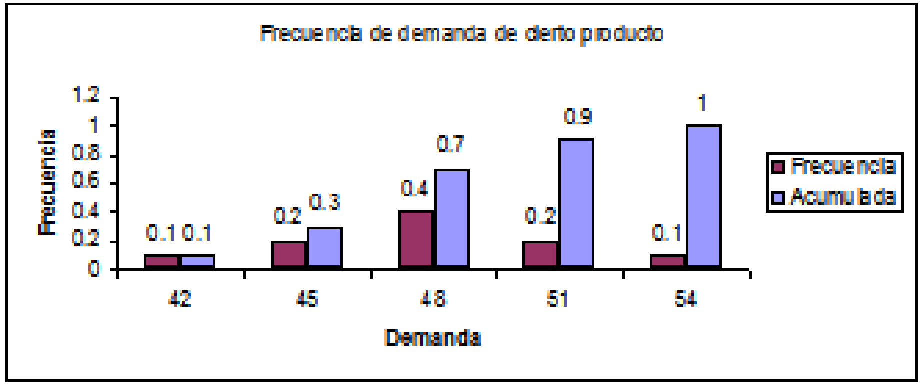


Frecuencia de demanda de cierto producto
Frecuencia
1.2
1
0.8
0.6
0.4
0.2
0
0.1 0.1
0.2
0.3
0.4
0.7
0.2
0.9
0.1
1
42
45
48
51
54
Demanda
Frecuencia
Acumulada

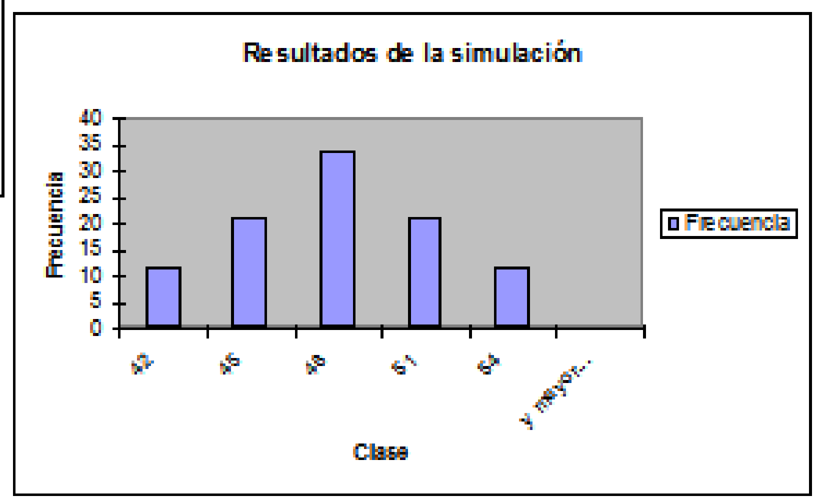
Resultados de la simulación - Excel

Demanda	Frecuencia	Acumulada
42	0.1	0.1
45	0.2	0.3
48	0.4	0.7
51	0.2	0.9
54	0.1	1

48	42	48	51
45	54	54	45
48	45	51	42
51	45	54	51
51	51	45	51
54	48	54	54
42	51	42	54
48	51	51	54
51	42	51	48
45	45	54	48
45	51	48	48
42	48	48	48
42	45	51	45
45	48	48	42
45	48	51	51
42	51	42	48
45	48	45	51
48	45	42	48
48	54	48	45
48	48	48	51
48	48	48	48
48	51	48	45
54	51	54	45
48	45	42	48
48	45	45	48



Resultados de la simulación			
Estadísticas		Clase	Frecuencia
		42	12
Media	48	45	21
Error típico	0.354196	48	34
Mediana	48	51	21
Moda	48	54	12
Desviación estándar	3.541956	y mayor...	0
Varianza de la muestra	12.54545		
Curtosis	-0.740163		
Coeficiente de asimetría	0		
Rango	12		
Mínimo	42		
Máximo	54		
Suma	4800		
Cuenta	100		



Ejemplo 2

Se quiere analizar la rentabilidad de una inversión a cuatro años de cierto producto.

Se estima que la demanda del mismo está uniformemente distribuida con valores entre 8 y 13 unidades anuales con un precio de $35,000 cada una. Los costos fijos anuales son de $15,000 y los variables del 75% de las ventas. La depreciación anual del equipo necesario es de $10,000 y se estima una inversión de $150,000. El costo de capital es del 10% y los impuestos se calculan en base a una tasa del 34%.

A fin de tener un estimado realista, se sugiere desarrollar 100 simulaciones del comportamiento de la inversión y calcular el promedio del valor presente de la misma antes de tomar una decisión.

	B	C	D	E	F
1	$ 150,000.00		Costos variables	75%	de lo
2	$ 35,000.00		Costo de capital	10%	
3	$ 15,000.00		Impuestos	34%	
4	$ 10,000.00		Demanda	max	13
5				min	8

	A	B	C
13	Unidades vendidas		=REDONDEAR($F$5+($F$4-$F$5)*ALEATORIO(),0)
14	Ingresos por ventas		=$B$2*C13
15	Costos variables		=$E$1*C14
16	Costos fijos		=$B$3
17	Depreciación anual		=$B$4
18	UAI		=C14-SUMA(C15:C17)
19	Impuestos		=$E$3*C18
20	Utilidad Neta		=C18-C19
21	Mas: Depreciación anual		=$B$4
22	Flujo Neto	=-B12	=C20+C21
23			
24		VPN	=VNA(E2,C22:F22)+B22

Inversión Inicial	$	150,000.00	Costos variables		75%	de los ingresos
Precio de venta	$	35,000.00	Costo de capital		10%	
Costos fijos	$	15,000.00	Impuestos		34%	
Depreciación anual	$	10,000.00	Demanda	max	13	
				min	8	

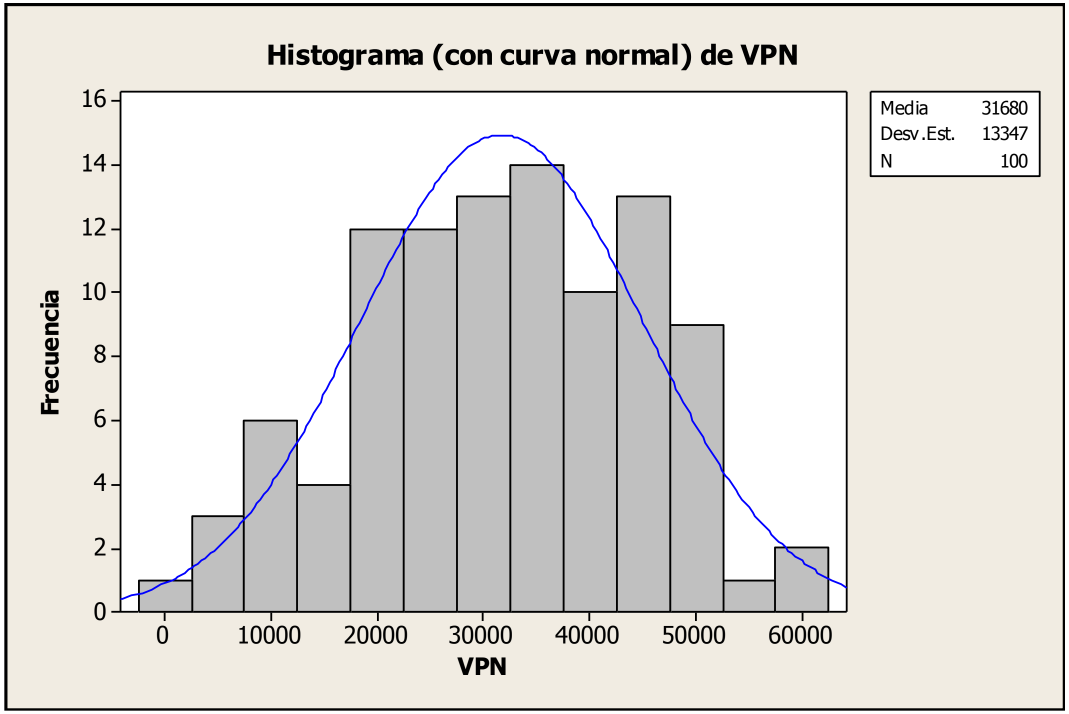
Año	0	1	2	3	4
Inversión Inicial	$ 150,000.00				
Unidades vendidas		10	10	13	10
Ingresos por ventas		$ 350,000.00	$ 350,000.00	$ 455,000.00	$ 350,000.00
Costos variables		$ 262,500.00	$ 262,500.00	$ 341,250.00	$ 262,500.00
Costos fijos		$ 15,000.00	$ 15,000.00	$ 15,000.00	$ 15,000.00
Depreciación anual		$ 10,000.00	$ 10,000.00	$ 10,000.00	$ 10,000.00
UAII		$ 62,500.00	$ 62,500.00	$ 88,750.00	$ 62,500.00
Impuestos		$ 21,250.00	$ 21,250.00	$ 30,175.00	$ 21,250.00
Utilidad Neta		$ 41,250.00	$ 41,250.00	$ 58,575.00	$ 41,250.00
Mas: Depreciación anual		$ 10,000.00	$ 10,000.00	$ 10,000.00	$ 10,000.00
Flujo Neto	$ (150,000.00)	$ 51,250.00	$ 51,250.00	$ 68,575.00	$ 51,250.00

Valor presente $ 25,472.13

	A	B	
1	Simulacion	VPN	
2	1	25,472.13	
3	2	25,073.75	
4	3	51,595.91	
5	4	13,761.20	
6	5	25,073.75	
7	6	31,550.46	
8	7	29,495.42	
9	8	50,807.03	
10	9	28,639.49	
11	10	44,338.21	
12	11	16,005.57	
13	12	50,018.15	
14	13	24,643.81	
15	14	33,412.22	
16	15	3,344.03	
17	16	20,738.85	
18	17	21,567.17	
19	18	38,567.55	
20	19	35,494.86	
21	20	36,883.29	
22	21	46,073.75	
23	22	49,106.99	

Resultados de la simulación

Variable	N	Media	Desv.Est.	Mínimo	Mediana	Máximo
VPN	100	31680	13347	2038	32454	59485



Resultados por MINITAB