



Informática y Cálculo Numérico

Máster en Ciencias Actuariales y Financieras
Curso 2023-2024

Prof. José Luis Vilar Zanón
Departamento de Economía Financiera y Actuarial y Estadística
Universidad Complutense de Madrid.



Contenido:

Tercera parte: Introducción al Cálculo Numérico:

1. Aproximación de raíces de ecuaciones: el método de Newton.
2. Aproximación de soluciones de ecuaciones diferenciales ordinarias: el método (la poligonal) de Euler.
3. Integración numérica: fórmulas de Newton Cotes.

Cuarta Parte: Simulación Monte Carlo

4. Generación de números aleatorios uniformes.
5. Métodos estándar para la generación de números aleatorios.
Generación de números aleatorios de v.a. estándar.
6. Método Monte Carlo de integración.

Bibliografía



1. Aproximación de raíces: el problema

Dada una ecuación $F(x) = 0$

se trata de calcular sus raíces. Por ejemplo:

$$F(x) = x^2 - 2x + 1 = 0 \quad F(\lambda) = -10 + \frac{5}{(1+\lambda)^2} + \frac{6}{(1+\lambda)^3} + \frac{9}{(1+\lambda)^5} = 0$$

Podemos estar en dos situaciones:

- La ecuación es resoluble simbólicamente, como en el caso 1

$$F(x) = x^2 - 2x + 1 = 0 \Leftrightarrow x = 1$$

- La ecuación NO es resoluble simbólicamente, como en el caso 2

Para estos casos son necesarias las técnicas numéricas para la aproximación de las raíces.



1. Aproximación de raíces: utilidad

La aproximación de raíces es muy útil. Por ejemplo:

- El cálculo de la tasa interna de retorno (TIR) de un flujo de caja
- El cálculo de estimadores máximo-verosímiles en modelos de probabilidad (estimación puntual)
- Cualquier situación en donde surja una ecuación o sistema de ecuaciones no resoluble numéricamente, por ejemplo dentro de un proceso de optimización, etc...



1. Aproximación de raíces: Método de Newton (1)

Vamos a introducir el método de Newton-Raphson

1. Tomamos una aproximación inicial o **semilla** x_0
2. La k -ésima iteración del método nos puede dar la aproximación:

$$x_k = x_{k-1} - \frac{F(x_{k-1})}{F'(x_{k-1})}, \quad k = 1, 2, \dots, \quad \lim_{k \rightarrow \infty} x_k = x^* \text{ raíz}$$

¿Qué significado gráfico tiene este método?

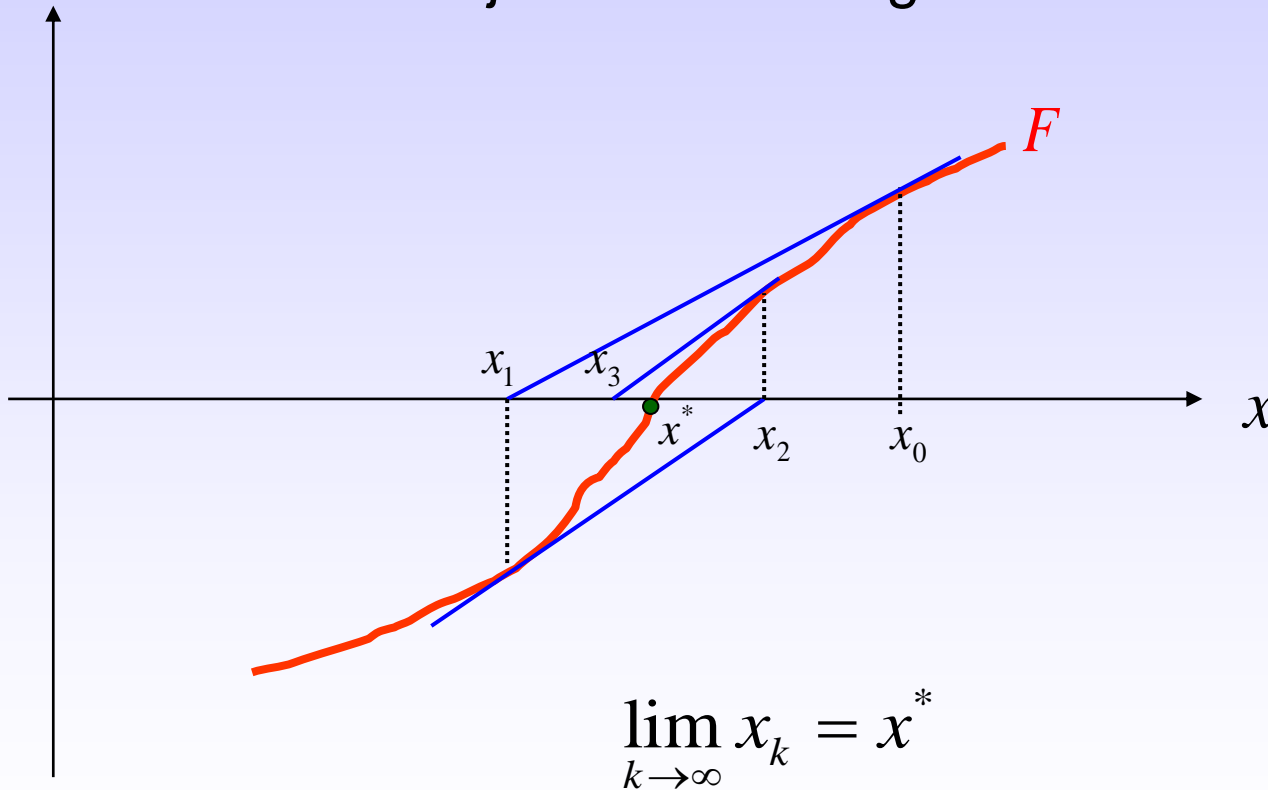
Primero tenemos que visualizar la ecuación y su raíz:

la raíz es el punto intersección de la gráfica de la función F con el eje de abscisas.



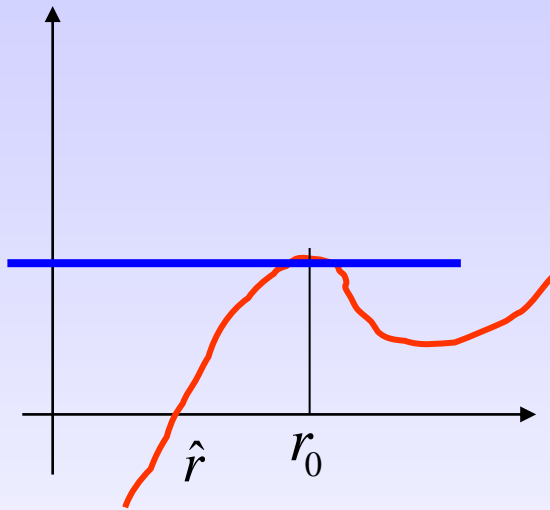
1. Aproximación de raíces: Método de Newton (2)

Gráficamente: Sustituimos F por su aproximación lineal (tangente) en x_k y la aproximación a la raíz de F es el corte con el eje de dicha tangente.

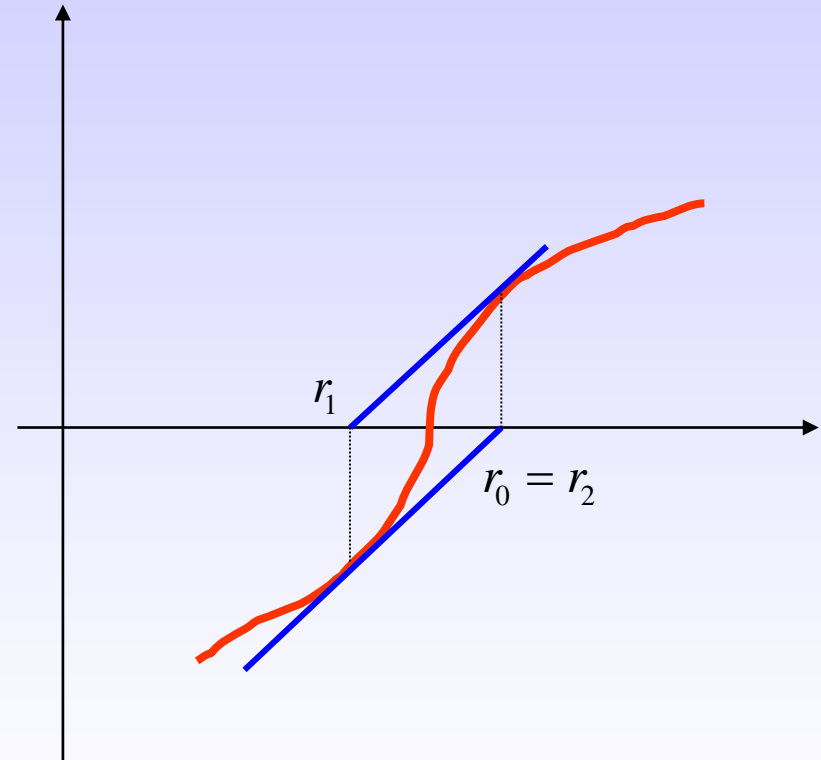




1. Aproximación de raíces: El método de Newton no siempre funciona...



No convergencia hacia la raíz



Ciclado



1. Aproximación de raíces: Ejercicios

Ejercicios nº1 al 3:

- Aproximación de raíces de ecuaciones y de sistemas en R.
- El paquete rootSolve
- Cálculo de la Tasa Interna de Retorno (TIR).
- Estimación Máximo Verosímil para la binomial negativa.



rootSolve



2. Aproximación de soluciones de EDO: definiciones previas

Necesitamos dar una serie de definiciones:

Ecuación Diferencial Ordinaria: es cualquier ecuación cuyas indeterminadas sean: un función $y(t)$, sus derivadas sucesivas y la variable independiente t :

$$f(t, y, y^{(1)}, \dots, y^{(n)}) = 0$$

Ejemplos:

$$y' = y \quad y' = 2t(y + 3) \quad y''' + 2y'' - 19y' - 20y = 0$$

$$y' = 3y^2 + t \quad y' = y^2 + t^2$$

Orden de una EDO: es el orden de la derivada máxima que aparece en la ecuación. (nosotros solo trabajamos en orden 1)

Las EDO's de orden 1 también se pueden expresar así:

$$y' = H(t, y)$$



2. Aproximación de soluciones de EDO: definiciones previas

Solución general de una EDO o curva integral: Es el conjunto de todas las funciones $y(t)$ que verifican idénticamente la EDO sobre un intervalo.

Condición inicial (CI): Es un valor al que se iguala la solución para un valor determinado de la variable independiente (que por defecto será el $t=0$). Por ejemplo: $y(0)=1$.

Solución Particular de una EDO: Es aquella función, perteneciente a la solución general, que verifica la condición inicial.

Problema de Valor Inicial (PVI): Un problema de valor inicial consiste en una EDO+CI. Si está “bien” planteado la sol. es ! . Por ejemplo en el caso de orden 1 que es el que vamos a utilizar:

$$\begin{cases} y' = H(t, y) \\ y(0) = y_0 \end{cases}$$



2. Aproximación de soluciones de EDO: el problema

Dado un PVI:

$$\begin{cases} y' = H(t, y) \\ y(0) = y_0 \end{cases}$$

se trata de encontrar la solución particular $y(t)$.

Pueden darse dos casos:

- La EDO es integrable simbólicamente: cambio de variable que la reduce a variables separadas+ cálculo de primitivas + aplicación de la CI para determinar la cte. de integración. Este es el caso de las EDO's anteriores de color negro.
- La EDO NO es integrable simbólicamente y tenemos que aplicar un método que aproxime numéricamente la solución sobre unos t 's previamente elegidos.

Este es el caso de las EDO's de color rojo, por ejemplo.



2. Ejemplo de aproximación de la solución de una EDO: el método de Euler (1).

- Consideremos el siguiente PVI
$$\begin{cases} y' = H(t, y) \\ y(0) = y_0 \end{cases}$$
- Podemos mirar a una EDO de 1º orden como una ecuación que relaciona el incremento $dy = y(t + dt) - y(t)$ con el incremento del tiempo t ($dt > 0$):

$$\frac{dy}{dt} = H(t, y) \xrightarrow[\text{Antes del paso al límite en la derivada}]{dt > 0} dy \simeq H(t, y)dt$$

Interpretación: La EDO nos dice como debe incrementarse (aproximadamente) la solución sobre un intervalo de tiempo arbitrariamente pequeño (amplitud $dt > 0$).



2. Aproximación de soluciones de EDO: la poligonal de Euler (2).

- Consideramos una red de puntos:

$$t_k = k \, dt, \quad k = 0, 1, 2, \dots$$

$$y(0) = y_0 \quad (CI)$$

$$y(t_1) = y_0 + dy \simeq y_0 + H(0, y_0)dt = y_1$$

$$y(t_2) = y(t_1) + dy \simeq y_1 + H(t_1, y_1)dt = y_2$$

$$\vdots \qquad \qquad \qquad \vdots$$

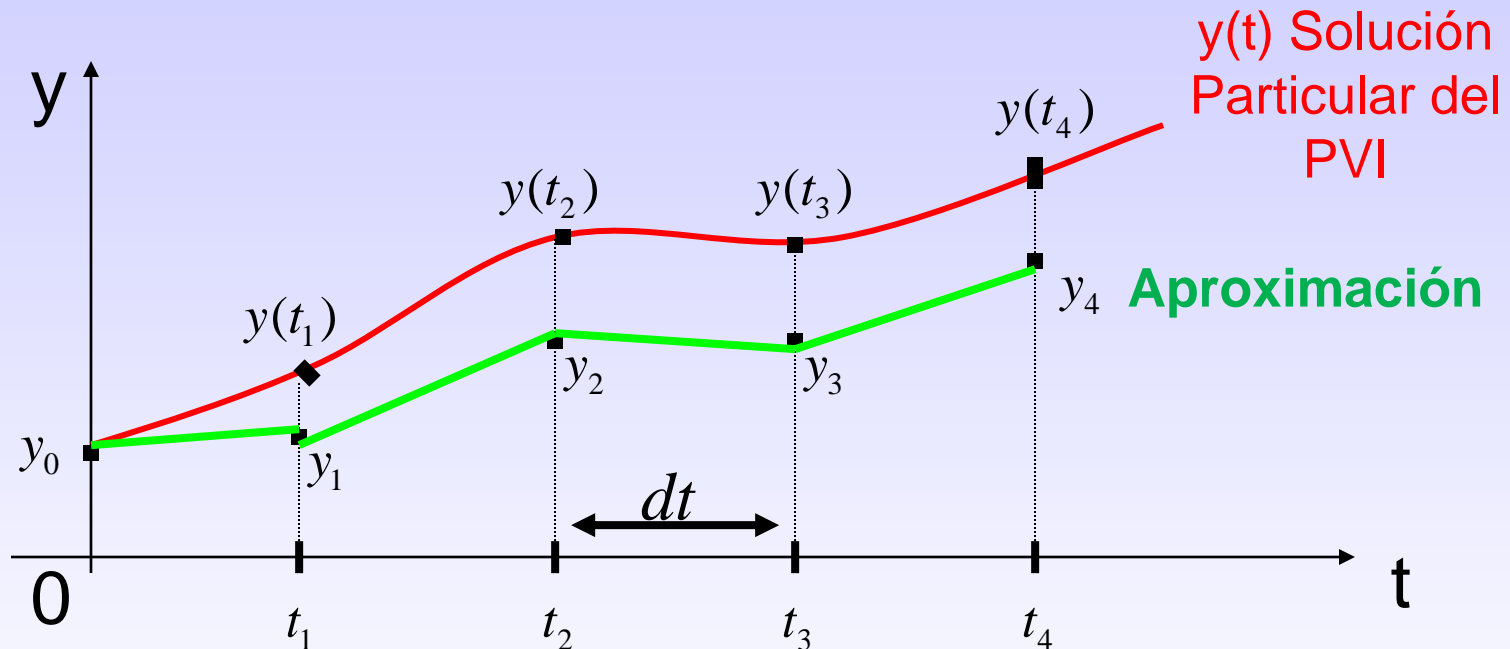
$$y(t_n) = y(t_{n-1}) + dy \simeq y_{n-1} + H(t_{n-1}, y_{n-1})dt = y_n$$

$$\vdots \qquad \qquad \qquad \vdots$$



2. Aproximación de soluciones de EDO: la poligonal de Euler (3).

- La poligonal de Euler $P_n(t)$ es la que une los puntos (t_n, y_n)



- Cuando $dt \searrow 0$, la poligonal se “confunde con” (tiende a), la solución particular del PVI: $\lim_{dt \searrow 0} P_n(t) = y(t)$



2. Aproximación de soluciones de EDO: Ejemplo

- PVI:
$$\begin{cases} y' = \alpha y \\ y(0) = y_0 \end{cases}$$

- La integración es sencilla:

$$\frac{dy}{dt} = \alpha y \Leftrightarrow \frac{dy}{y} = \alpha dt \Leftrightarrow \int \frac{dy}{y} = \alpha \int dt \Leftrightarrow$$

$$\Leftrightarrow \log(y) = \alpha t + K \Leftrightarrow y(t) = Ke^{\alpha t}, \quad \forall K \in \mathbb{R}$$

Solución General



2. ¿Como se calcula una solución particular de la ecuación?

- Imponiendo la CI:

$$y(0) = Ke^{\alpha 0} = K = y_0$$

- **Una Solución particular es**

$$y(t) = y_0 e^{\alpha t}$$



2. Como se aproxima la solución de la EDO

- Interpretamos la EDO en el sentido de los incrementos:

$$dy = \alpha y dt \quad (dt > 0)$$

- Y aplicamos lo dicho anteriormente. Algunos ejemplos **R** resueltos mediante el paquete **deSolve**:



Resolucion de
DO paquete deSolv



2. Aproximación de soluciones de EDO: Ejercicio

Ejercicio nº4:

- Aproximación de soluciones de PVI utilizando la función **ode()** del paquete **deSolve**
- Representación gráfica de las aproximaciones:
 - **plot()**, **par(new=TRUE)**, **curve()**

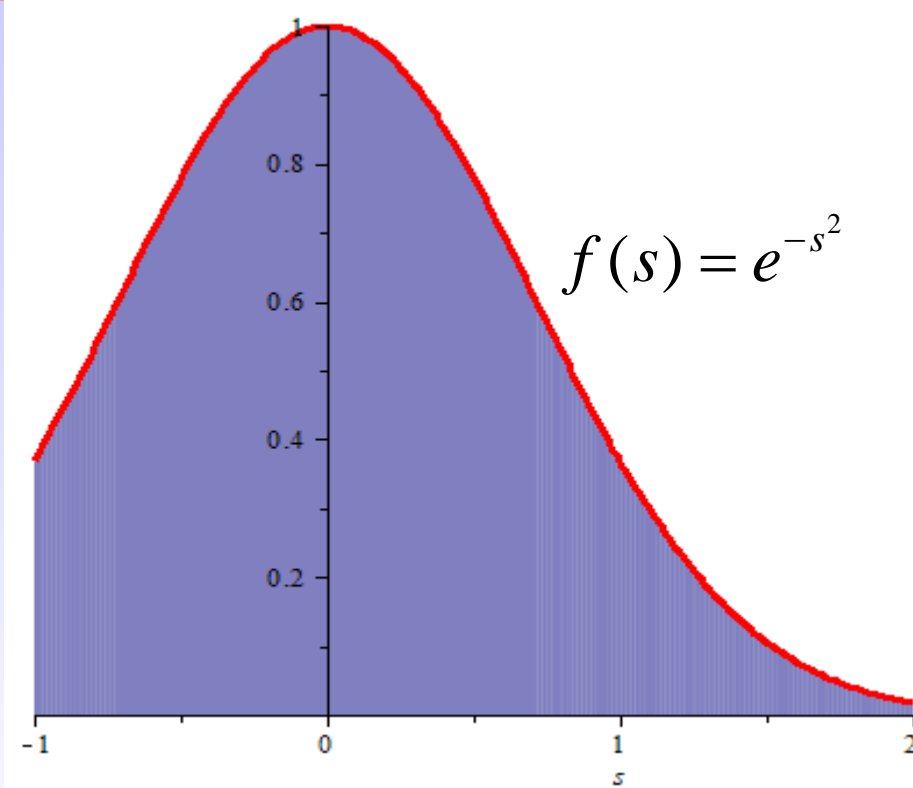


3. Integración numérica: El problema

Se trata de calcular:

$$f : [a, b] \rightarrow \mathbb{R}, \quad I(f) = \int_a^b f(s) ds$$

(Es posible que alguno de los límites de integración, o ambos, sean ∞).



El cálculo de integrales es una de las herramientas más importantes en seguros y finanzas: cálculo de probabilidades, tarificación, valoración de activos, medición de riesgos, etc, etc... dependen de su correcto cálculo.



3. Integración numérica: El problema

Los Teoremas Fundamentales del Cálculo garantizan:

•1º TFC: Si f es una función continua entonces la Función

$$F(x) = \int_a^x f(s)ds, \forall x \in [a, b]$$

es derivable y $F'(x) = f(x), \forall x \in [a, b]$ (F es una primitiva)

•2ºTFC: Si F es una primitiva de f entonces:

$$\int_a^b f(s)ds = F(b) - F(a) \quad (\text{Regla de Barrow})$$

Este proceso es lo que denominamos una integración simbólica. El problema es que casi nunca se puede llevar a cabo ya que , en general, ¡ F no será expresable en términos de funciones elementales! (e.d., desconocemos su expresión)



3. Integración numérica: una idea general

Por tanto, es necesario tener métodos numéricos que aproximen el valor de la integral. Esto es la Integración Numérica.

Para calcular $I(f) = \int_a^b f(x)dx$, debemos encontrar una

familia $\{f_n(x) : n \geq 1\}$ que se “acerque” bien a $f(x)$ y calcular

$$I_n(f) = \int_a^b f_n(x)dx = I(f_n) \approx I(f)$$

El **error absoluto** cometido al emplear la función f_n es el V.Abs. de la diferencia entre la integral exacta y su aproximación:

$$\begin{aligned} E(f_n) &= \left| \int_a^b f(x)dx - \int_a^b f_n(x)dx \right| \\ &= \left| \int_a^b (f(x) - f_n(x))dx \right| \leq \int_a^b |f(x) - f_n(x)|dx \end{aligned}$$



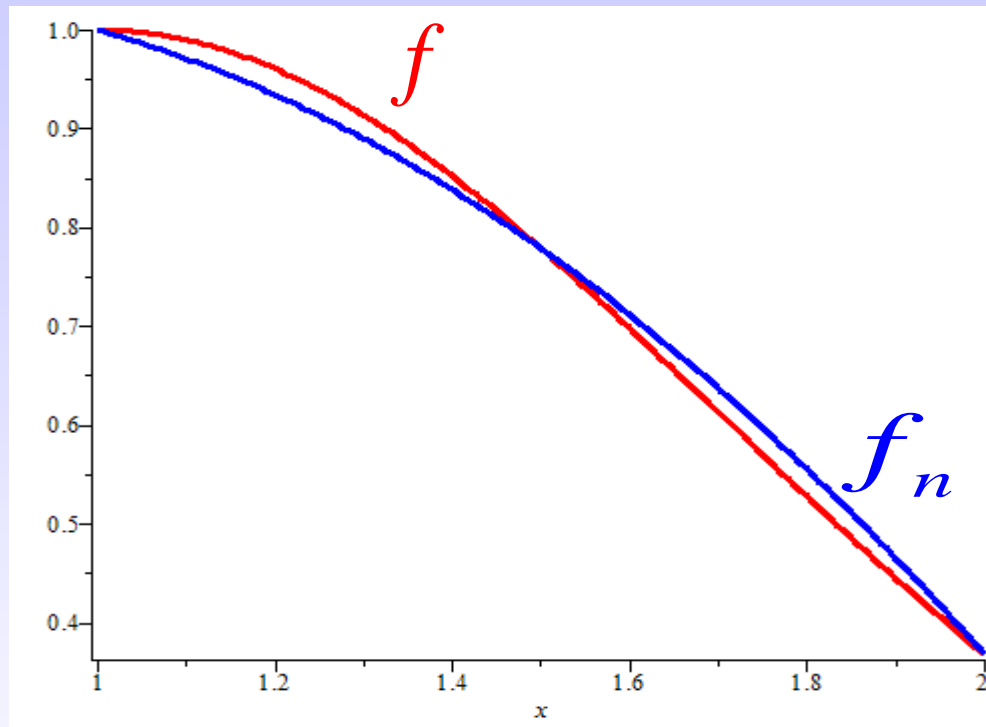
3. Integración numérica: una idea general

Entonces para que el método sea “Bueno”, su error tiene que tender a 0 a medida que $f_n \rightarrow f$

$$E_n(f) \leq \int_a^b |f(x) - f_n(x)| dx \leq (b - a) \max_{x \in [a, b]} (f(x) - f_n(x)) \xrightarrow{n \rightarrow \infty} 0$$



3. Integración numérica: una idea general



Muchos métodos numéricos toman como f_n a un polinomio que interpola a f en n puntos x_{jn} $j=1 \dots n$. La aproximación se expresa:

$$I_n(f) = \sum_{j=1}^n w_{jn} f(x_{jn}), \quad n \geq 1, \begin{cases} w_{jn} & \text{son los } \textit{pesos} \\ x_{jn} & \text{son los } \textit{nodos} \end{cases}$$

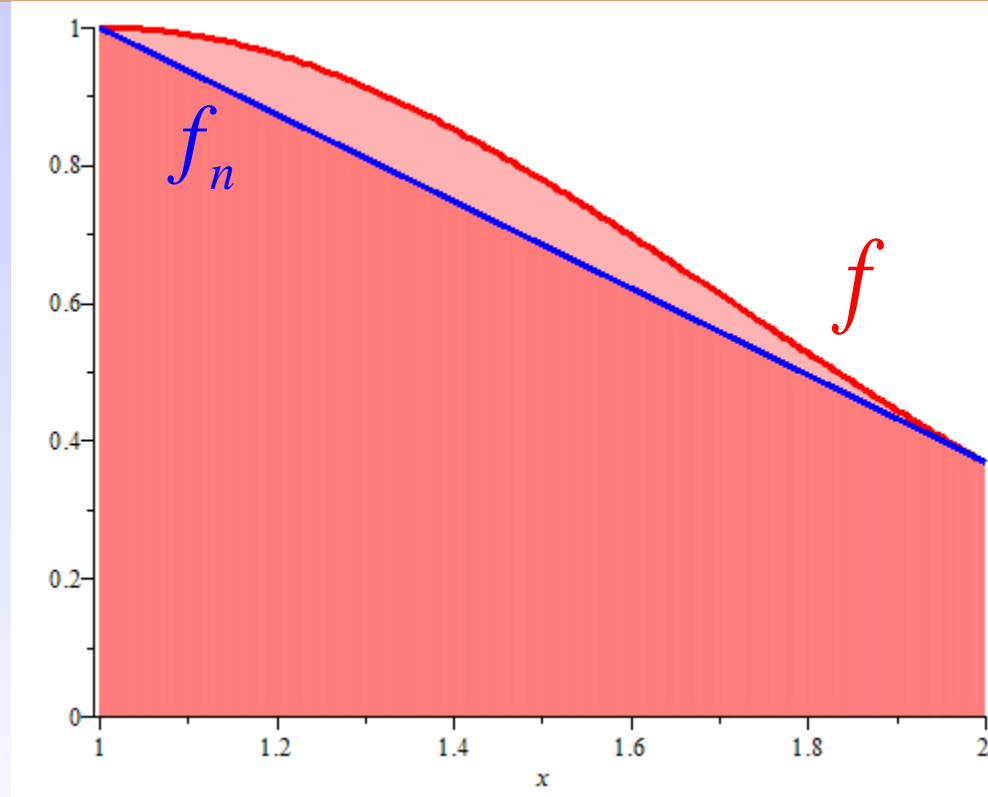


3. La regla del trapecio

Se interpola un polinomio de grado 1. $b - a = h$

$$I_1(f) = \frac{h}{2} (f(a) + f(b))$$

$$E_1(f) = \frac{-h^3}{12} f''(\xi), \quad \xi \in [a, b]$$



h tiene que ser pequeño. Para ganar precisión la regla puede ser aplicada descomponiendo $[a, b]$ en subintervalos . Se trata entonces de una **regla compuesta**.



3. La regla de Simpson

Mejoramos la aproximación interpolando un polinomio $p_2(x)$ de grado 2 en los puntos $a, b, c = (a + b)/2$

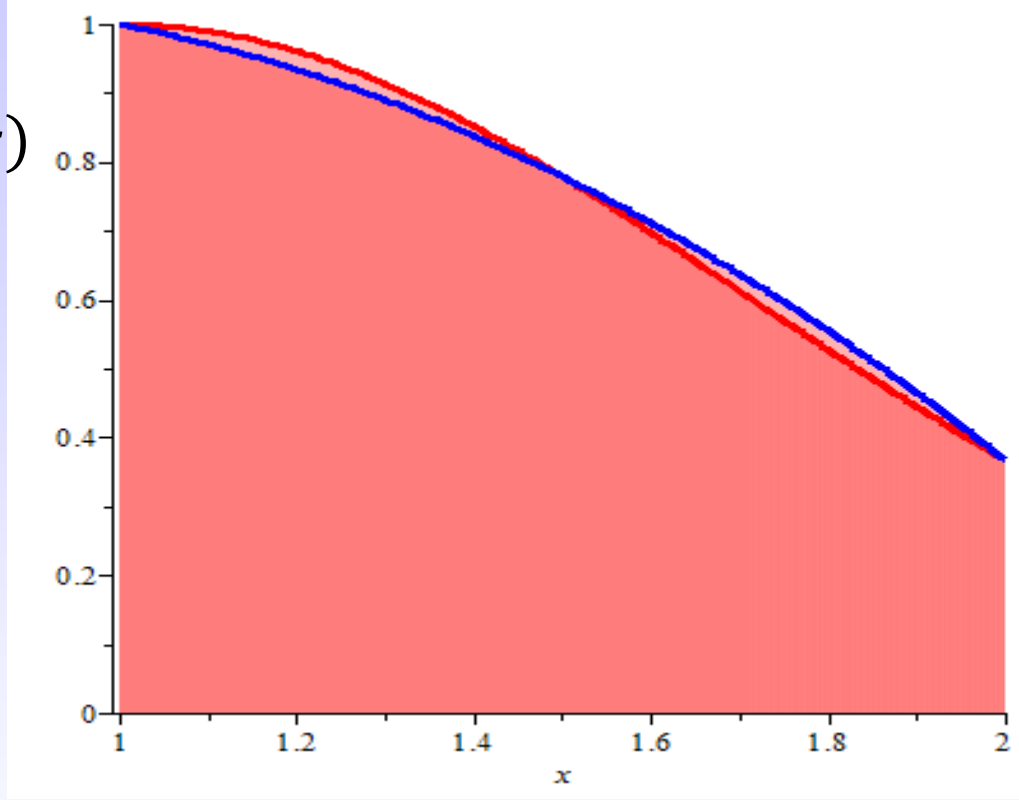
$$I_2(f) = \frac{h}{3} (f(a) + 4f(c) + f(b))$$

$$E_2(f) = \frac{-h^5}{90} f^{(4)}(\xi), \quad \xi \in [a, b]$$



El error es nulo cuando f es un polinomio de grado ≤ 3

$$p_2(x) = \frac{(x-c)(x-b)}{(a-c)(a-b)} f(a) + \frac{(x-a)(x-b)}{(c-a)(c-b)} f(c) + \frac{(x-a)(x-c)}{(b-a)(b-c)} f(b)$$





3. Fórmulas de Newton-Cotes

$$\forall n \geq 1, \quad h = \frac{b-a}{n}, \quad x_j = a + jh \quad (j = 0, 1, \dots, n)$$

Tomamos el polinomio que interpola a: $f(x_0), \dots, f(x_n)$

$$I(f) = \int_a^b f(x)dx \approx I_n(f) = \int_a^b p_n(x)dx$$

$$n = 3: \quad I_3(f) = \frac{3h}{8} (f(a) + 3f(a+h) + 3f(b-h) + f(b)), \quad E_3(f) = \frac{-3h^5}{80} f^{(4)}(\xi)$$

$$n = 4: \quad I_4(f) = \frac{2h}{45} \left(7f(a) + 32f(a+h) + 12f\left(\frac{a+b}{2}\right) + 32f(b-h) + 7f(b) \right),$$

$$E_4(f) = \frac{8h^7}{945} f^{(6)}(\xi)$$



3. Grado de Precisión de una fórmula de integración numérica

Una fórmula de integración numérica $\tilde{I}(f)$ que aproxima $I(f)$ tiene grado de precisión m si

$$\tilde{I}(f) = I(f) \quad \forall \text{polinomio } f(x) \text{ de grado } \leq m$$

$$\tilde{I}(f) \neq I(f) \quad \text{para algún polinomio } f(x) \text{ de grado } m+1$$

Ejemplo:

- Para $n = 1, 3$ los grados de precisión de N-C son 1 y 3
- Para $n = 2, 4$ los grados de precisión de N-C son 3 y 5

Las fórmulas de N-C para n par aportan un grado de precisión adicional en comparación con las mismas fórmulas para n impar.



3. Integración numérica sobre un intervalo no acotado

La idea aquí consiste en “embutir” todo el intervalo no acotado de integración dentro de uno acotado y aplicar entonces lo anterior. Pero es un paso delicado y hay que leer bien la ayuda de las funciones de integración numérica que se usen.

Recordemos como es el cambio de variable bajo el signo integral:

$$x = x(s) \text{ monótona, } (x^{-1}(a) = s_1, x^{-1}(b) = s_2), \quad dx = x'(s)ds.$$

Entonces:

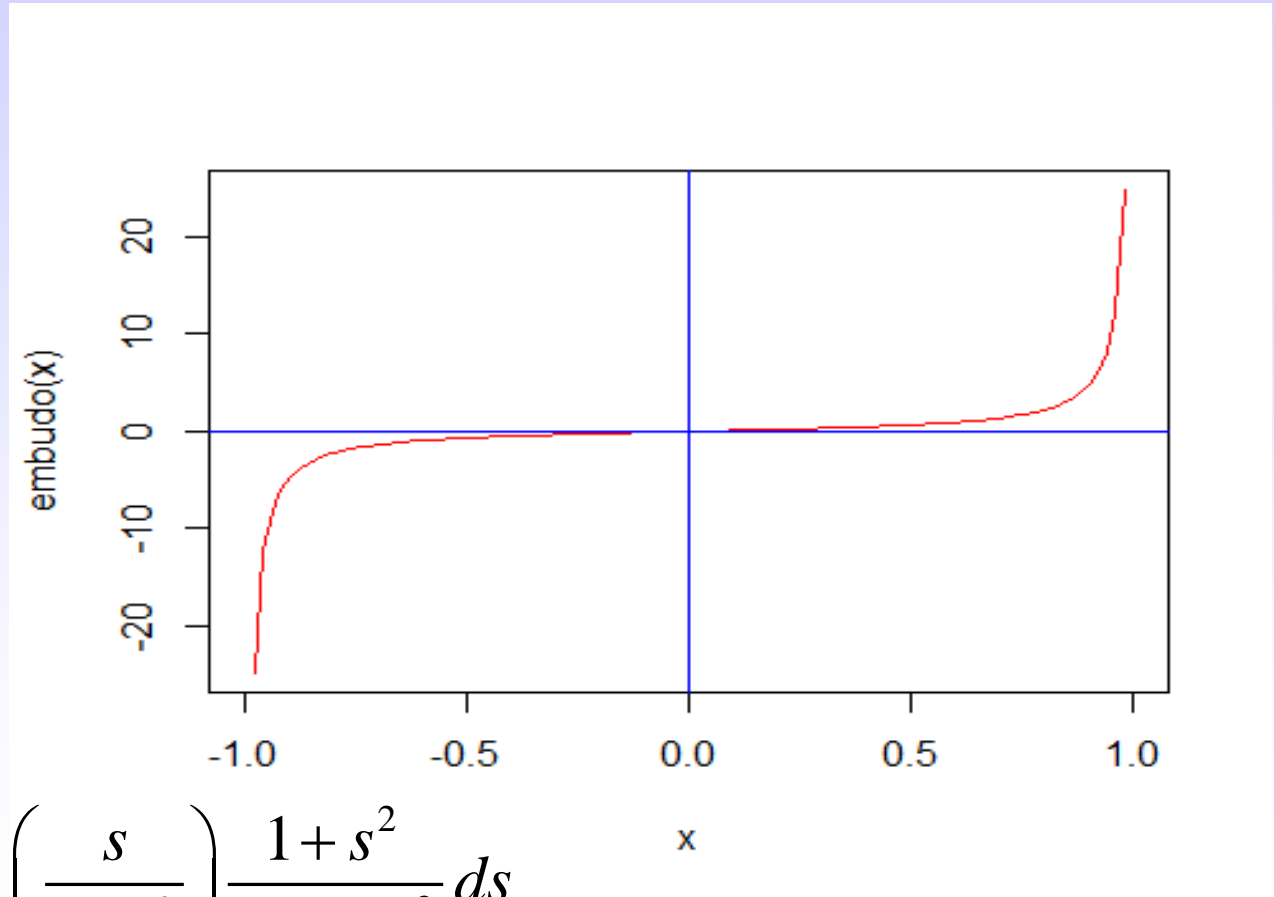
$$\int_a^b f(x)dx = \int_{s_1}^{s_2} f(x(s))|x'(s)|ds$$



3. Integración numérica sobre un intervalo no acotado

Si integramos sobre un intervalo infinito, podemos realizar un cambio que lo lleve a un intervalo acotado. Por ejemplo:

$$x(s) = \frac{s}{1-s^2}$$



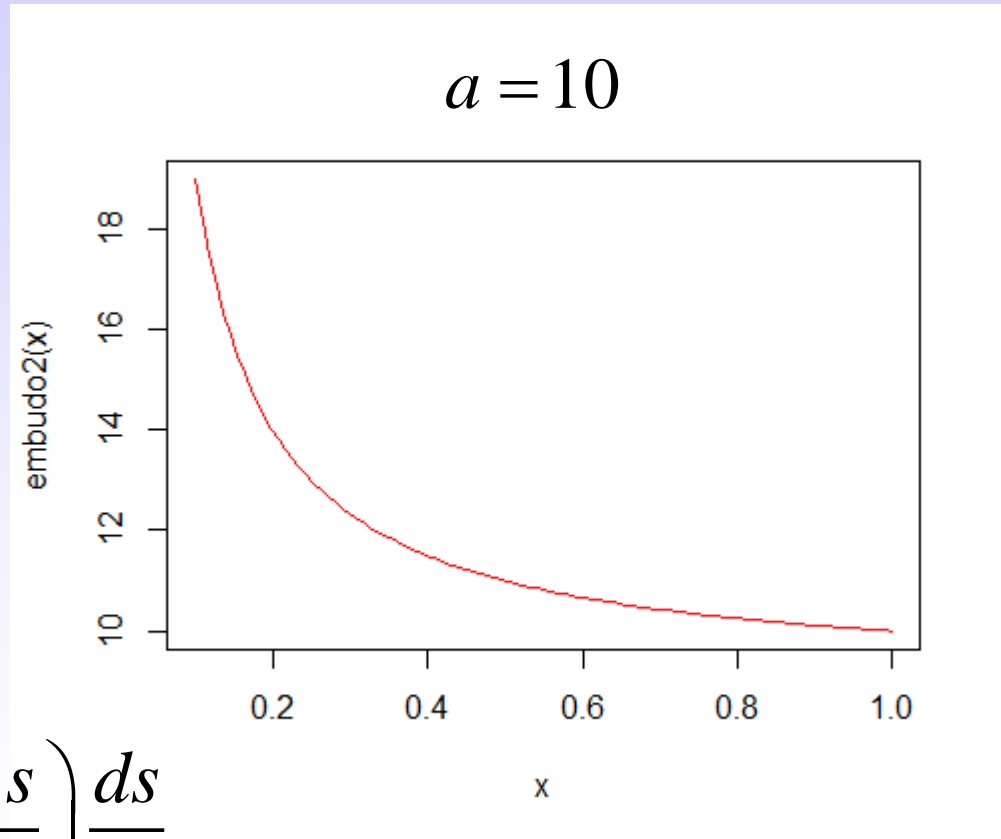
$$\int_{-\infty}^{+\infty} f(x) dx = \int_{-1}^1 f\left(\frac{s}{1-s^2}\right) \frac{1+s^2}{(1-s^2)^2} ds$$



3. Integración numérica sobre un intervalo no acotado

Si integramos sobre un intervalo infinito, podemos realizar un cambio que lo lleve a un intervalo acotado. Por ejemplo:

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



$$\int_a^{+\infty} f(x) dx = \int_0^1 f\left(a + \frac{1-s}{s}\right) \frac{ds}{s^2}$$

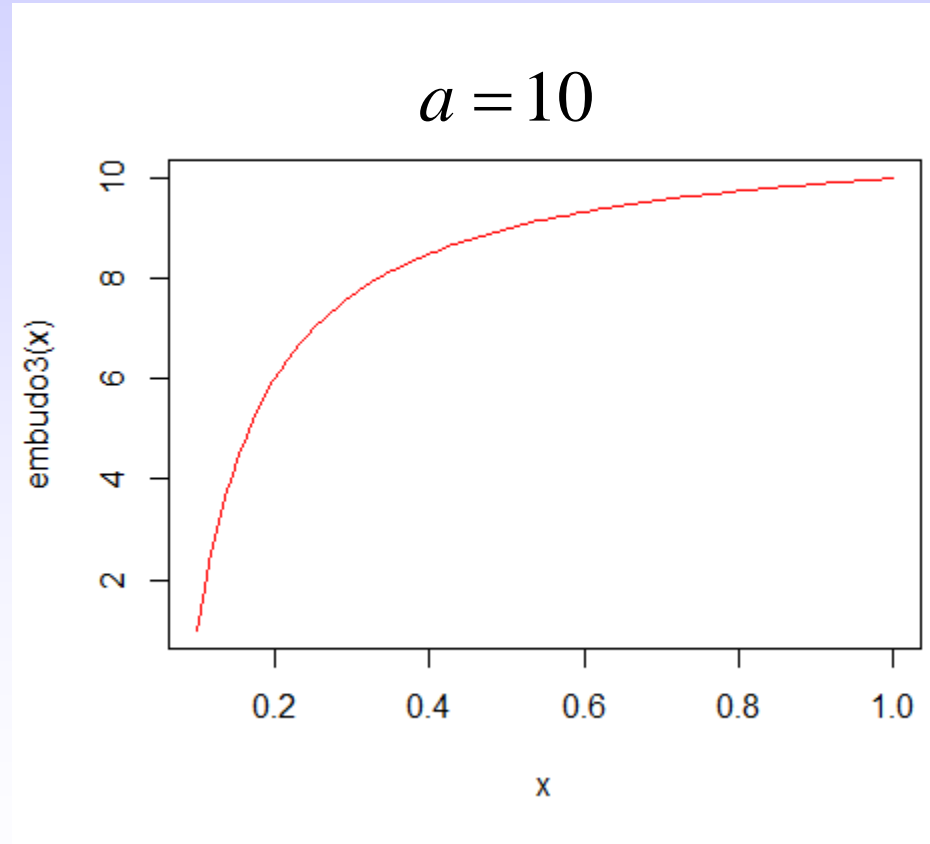


3. Integración numérica sobre un intervalo no acotado

Si integramos sobre un intervalo infinito, podemos realizar un cambio que lo lleve a un intervalo acotado. Por ejemplo:

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

$$\int_{-\infty}^a f(x) dx = \int_0^1 f\left(a - \frac{1-s}{s}\right) \frac{ds}{s^2}$$





Ejercicios

Ejercicio 5:

- Integración numérica en **R** con la función **integrate()** en una variable y
- paquete **cubature** , función **adaptIntegrate()** para funciones reales de varias variables.



integrate y pack
cubature



3. El problema de la alta dimensionalidad

Las integrales múltiples encierran un **problema** aparentemente **insoluble** desde el punto de vista del Cálculo Numérico:

la “**maldición de la dimensionalidad**”.

Para hacernos una idea, si para obtener un error pequeño se necesitase una fórmula de integración con 1000 nodos, en dimensión 10 la integral múltiple, ¡necesitaría de 10^{1000} nodos!

Esto hace que las fórmulas de integración numérica llevadas a dimensión grandes no sean operativas.



3. El problema de la alta dimensionalidad

Ejemplo: imaginemos los incrementos semanales del precio de un activo modelizados mediante:

$$\left(X_i \right)_{i=1}^{\infty} \text{ v.a.i.i.d. absolut. cont., } X \sim F(x) = P \{ X \leq x \}$$

¿Cómo calcular la probabilidad de que dentro de 10 semanas el precio sea $\geq z$?

$$¿ P \left\{ \sum_{i=1}^{10} X_i \geq z \right\} ?$$

¡Esto envolvería el cálculo de una una integral en dimensión 10!



3. Integración numérica y Método Monte Carlo.

- El método Monte Carlo es una herramienta para resolver el anterior problema, el problema de la alta dimensionalidad.

- El método depende de nuestra capacidad para simular muestras de los modelos de probabilidad.

De ahí la necesidad de aprender como se simula y lo que esto realmente significa.

- El desarrollo de los métodos de simulación corre parejo al desarrollo de las computadoras
- Un primer hito se establece en los años 40 del siglo XX
Una segunda revolución sucede al final de los años 90 debido a los altos estándares de calidad requeridos en las aplicaciones industriales, financieras, seguros etc.



3. Integración numérica y Método Monte Carlo.

Origen del nombre “Monte Carlo”: 1940.

- Dos matemáticos y un físico: **John Von Neumann, Stanislaw Ulam, Nicholas Metrópolis.**
- El Proyecto Manhattan, los Alamos National Laboratory (USA)
- Los documentos de trabajo se nombraban en clave.
- Al tío de Ulam le gustaba mucho ir a jugar al casino de Monte Carlo
- A uno de los documentos secretos, en donde se usaba por primera vez el método, le dieron el nombre clave “**Monte Carlo**”.



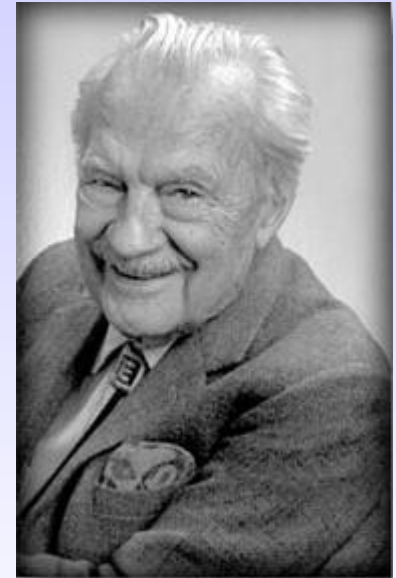
3. Integración numérica y Método Monte Carlo.



John Von Neumann
(Dic 28, 1903- Feb 8,1957)



Stanislaw Ulam
Apr 13, 1909 – May 13, 1984



Nicholas Metr6polis
Jun 11, 1915 – Oct 17, 1999



3. Integración numérica y Método Monte Carlo.

Método Monte Carlo= Simulación
+ Estimación
+Reducción de varianza

Además: la simulación de números aleatorios nos permite crear muestras de los modelos del mundo real para poder estudiar como se comportan dichos modelos. Es una poderosa herramienta de análisis del futuro.

Aplicaciones a seguros y finanzas: simulación de la siniestralidad o del precio de los activos, tarificación, valoración de activos, medición del riesgo, etc.... son análisis en donde la simulación desempeña un papel determinante cuando las técnicas simbólicas o analíticas no son satisfactorias.



4. Generación de números aleatorios uniformes

Las secuencias de números aleatorios $R_i \sim U(0,1)$

La materia prima de cualquier simulación:

Números aleatorios independientes $R_1, \dots, R_i, \dots \sim U(0,1)$

Problema: ¿Cómo generar una tal secuencia de números?

Una posible solución sería mediante mecanismos analógicos como por ejemplo:

- Lotería, ruleta, urna, etc...

Pero los inconvenientes son serios:

- Dichos dispositivos no interactúan fácilmente con los ordenadores
- Existe la necesidad de volver a usar la **misma corriente** de números aleatorios (p.ej. para controlar un código).



4. Generación de números aleatorios uniformes

Las secuencias de números pseudo-aleatorios R_i

Un posible remedio: Almacenamiento de la secuencia en algún periférico (CD rom, etc...). Pero esto ralentizaría demasiado la simulación (velocidad de acceso).

Solución definitiva: Se generan números **pseudo-aleatorios mediante reglas determinísticas.**

Ventajas:

- Generación muy rápida
- Se elimina el problema de la reproducibilidad de la secuencia
- Se elimina el problema del almacenamiento.

PRECAUCIÓN:

¡¡La regla determinística tiene que tener una fuerte apariencia aleatoria!! ;)



4. Generación de números aleatorios uniformes

Generadores Congruentes Lineales (GCL)

Una idea sencilla:

$$\begin{array}{l} \text{Dividendo} \\ \hline \text{Resto} \end{array} \left| \begin{array}{l} \text{Divisor} \\ \hline \text{Cociente} \end{array} \right. \longrightarrow \text{Resto} = \text{Dividendo} \bmod \text{Divisor}$$

Ejemplos:

- Divisor=2: solo hay **dos** puntos: $\{0,1\}$

$$0 = 2 \bmod 2, \quad 1 = 3 \bmod 2, \quad 0 = 4 \bmod 2, \quad \dots$$

- Divisor=3: solo hay **tres** puntos: $\{0,1,2\}$

$$0 = 3 \bmod 3, \quad 1 = 4 \bmod 3, \quad 2 = 5 \bmod 3, \quad 0 = 6 \bmod 3, \quad \dots$$

- Ejercicio: ¿Qué pasa en mod 4?

- Ejercicio: ¿Qué pasa en mod 10^3 ?



4. Generación de números aleatorios uniformes

Generadores Congruentes Lineales (GCL)

Una idea sencilla:

$$\begin{array}{l} \text{Dividendo} \\ \hline \text{Resto} \end{array} \left| \begin{array}{l} \text{Divisor} \\ \hline \text{Cociente} \end{array} \right. \longrightarrow \text{Resto} = \text{Dividendo} \bmod \text{Divisor}$$

Secuencia de enteros no negativos $\{X_i : i = 1, 2, \dots\}$

$$X_i = (aX_{i-1} + c) \bmod m \in [0, m-1] \rightarrow R_i = \frac{X_i}{m} \in [0, 1]$$

Los parámetros son:

- El multiplicador: $a > 0$
 - La semilla: $X_0 \geq 0$
 - El incremento: $c \geq 0$
- $$\left. \begin{array}{l} \\ \\ \\ \end{array} \right\} \in [0, m-1]$$
- El módulo: $m > 0$, se tomara como $m = 2^q$

La idea es que si m es suficientemente grande los R 's estén tan cerca unos de otros que puedan ser considerados v.a continua.



4. Generación de números aleatorios uniformes

Generadores Congruentes Lineales (GCL)

El periodo de un GCL: $\lambda = \min \{n \in \mathbb{N} : X_n = X_0\}$

- El periodo máximo posible es el módulo: m
- Luego cuanto más grande sea m , mejor.

¿ Cuando se obtiene el periodo $\lambda = m$ máximo?

$$\lambda = m \Leftrightarrow \begin{cases} c, m \text{ son primos relativos } (mcd(c, m) = 1) \\ a - 1 \text{ es multiplo de } q, \text{ } q \text{ factor primo de } m \\ a - 1 \text{ es multiplo de } 4 \text{ si } m \text{ lo es} \end{cases} \quad (1.1)$$

Clasificación de los GCL:

- Mixtos: $c > 0$
- Multiplicativos: $c = 0$



4. Generación de números aleatorios uniformes

Generadores Congruentes Lineales (GCL)



generadores
congruentes lineales

Ejemplo 1:

$$X_i = (9X_{i-1} + 3) \bmod 2^4 \in [0, 2^4 - 1], i = 1, 2, \dots, X_0 \in [0, 15]$$

Ejemplo 2: Una buena elección es $m = 2^b$ siendo b el número máximo de bits necesario para representar un entero. En muchas máquinas b=32, con un bit reservado para el signo. Se pueden representar entonces los enteros $[-2^{31}, 2^{31} - 1]$.

Tomando $m = 2^b$ el GCL tendrá periodo máximo cuando:

c sea impar, a-1 sea múltiplo de 4 (condiciones **(1.1)**)

$$X_i = (906185749X_{i-1} + 1) \bmod 2^{31} \in [0, 2^{31} - 1], i = 1, 2, \dots, X_0 \in [0, 2^{31} - 1]$$

Periodo de este GCL: $\lambda = 2^{31} = 2147483648$



4. Generación de números aleatorios uniformes

Tests de uniformidad para números pseudo-aleatorios

Nuestras secuencias de números pseudo-aleatorios tendrían que ser lo más parecidas posible a una sucesión de v.a.i.

$$R_i \sim U(0,1).$$

¿En qué medida es esto cierto?



validez gcl



4. Generación de números aleatorios uniformes

Ejercicios

Ejercicios:

6: Generadores Lineales Congruentes

7, 8, 9: ¿Cómo de bien nos engañan los Generadores
Congruentes Lineales?



4. Generación de números aleatorios uniformes

Generadores lineales de números en F_2

Una idea distinta y posterior para generar secuencias de números pseudo aleatorios es:

- Dado que los ordenadores trabajan en aritmética binaria (base dos), parece natural construir generadores de números por medio de recurrencias que usen la aritmética propia de los ordenadores.



4. Generación de números aleatorios uniformes

Generadores lineales de números en F_2

$$F_2 = (\{0,1\}, Xor, And) = (\{0,1\}, \oplus, \cdot)$$

INPUT		OUTPUT
A	B	A XOR B
0	0	0
0	1	1
1	0	1
1	1	0

INPUT		OUTPUT
A	B	A AND B
0	0	0
0	1	0
1	0	0
1	1	1

F_2 verifica una serie de propiedades algebraicas que se resumen diciendo que es un CUERPO:

- $(\{0,1\}, \oplus)$ es un grupo abeliano
- $(\{0,1\}, \cdot)$ es semigrupo (\cdot es asociativa)



4. Generación de números aleatorios uniformes

Generadores lineales de números en F_2

La recurrencia lineal mediante la cual se definen estos generadores es :

Comenzando con una semilla x_0 :

$$\left. \begin{array}{l} A \in M_{k \times k}(F_2) \text{ matriz de 0s y 1s} \\ x_n = (x_{n,0}, \dots, x_{n,k-1}) \text{ vector de k 0s y 1s} \end{array} \right\} \Rightarrow x_n = Ax_{n-1} \text{ ,}$$

$$B \in M_{\omega \times k}(F_2) \text{ matriz de 0s y 1} \Rightarrow y_n = Bx_n$$

$y_n = (y_{n,0}, \dots, y_{n,\omega-1})$ es un vector de ω 0s y 1s con el que se forma el n-esimo numero u_n de la secuencia (en base 2):

$$u_n = \sum_{l=1}^{\omega} y_{n,l-1} 2^{-l} \in (0,1) \text{ (Esta formado por } \omega \text{ bits)}$$



4. Generación de números aleatorios uniformes

Generadores lineales de números en F_2

El generador Mersenne-Twister se obtiene para una cierta elección de las matrices A , B y de las dimensiones n, k, ω .

- Su periodo es de $2^{19937} - 1$
- Respeta la distribución uniforme hasta dimensión 623 (¡!)
- Estas características hacen que sea el generador lineal de números aleatorios más usado actualmente.



4. Generación de números aleatorios uniformes

Generadores lineales de números en F_2

En R se pueden elegir muchos generadores lineales aunque por defecto toma el MT, y es mejor no cambiarlo.



generadores
aleatorios uniformes

Ejercicio nº 10: Generación de números aleatorios en R



5. Métodos estándar de generación de n° aleatorios: Método de Inversión de la F. de Distribución.

Tomemos dos v.a.: $X \sim F(x) = P\{X \leq x\}$
 $R \sim U(0,1), P\{R \leq r\} = r, 0 \leq r \leq 1$

Entonces: $F^{-1}(R) \sim F$

$$P\{F^{-1}(R) \leq x\} \stackrel{\uparrow}{=} P\{R \leq F(x)\} = F(x)$$

F monótona no decreciente

Así pues para simular los X 's pseudo-distribuidos según F :

1-Simular $R_i \sim U(0,1)$

2-Calcular $X_i = F^{-1}(R_i)$



5. Métodos estándar de generación de n° aleatorios: Método de Inversión de la F. de Distribución.

Ejemplo 3: Distribución exponencial $\text{Exp}(\lambda)$:

$$R = F(x) = 1 - e^{-\lambda x}, x \geq 0, \lambda > 0. \quad x = -\frac{1}{\lambda} \log(1 - R), R \in (0, 1)$$

Ejemplo 4: Distribución Weibull $\text{Wei}(b, c)$:

$$R = F(x) = 1 - e^{-\left(\frac{x}{b}\right)^c}, x \geq 0 \Leftrightarrow x = b \left(-\log(1 - R)\right)^{\frac{1}{c}}, R \in (0, 1)$$

Ejemplo 5: Distribución Logística $\text{Logi}(a, b)$:

$$R = F(x) = \left(1 + e^{-\frac{x-a}{b}}\right)^{-1}, x \in \mathbb{R}, a \in \mathbb{R}, b > 0 \Leftrightarrow x = a - b \log\left(\frac{1}{R} - 1\right), R \in (0, 1).$$

Ejemplo 6: Distribución de Cauchy $C(a, b)$:

$$R = F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x-a}{b}\right), x \in \mathbb{R}, a \in \mathbb{R}, b > 0 \Leftrightarrow x = a + b \tan\left(\pi\left(R - \frac{1}{2}\right)\right).$$



5. Métodos estándar de generación de n° aleatorios. Método del rechazo.

Queremos generar números a partir de una densidad

$$f(x) = \frac{h(x)}{\int_{-\infty}^{+\infty} h(x) dx}$$

Sin que dispongamos de una fórmula cerrada para F^{-1} .

Procedemos mediante los siguientes pasos

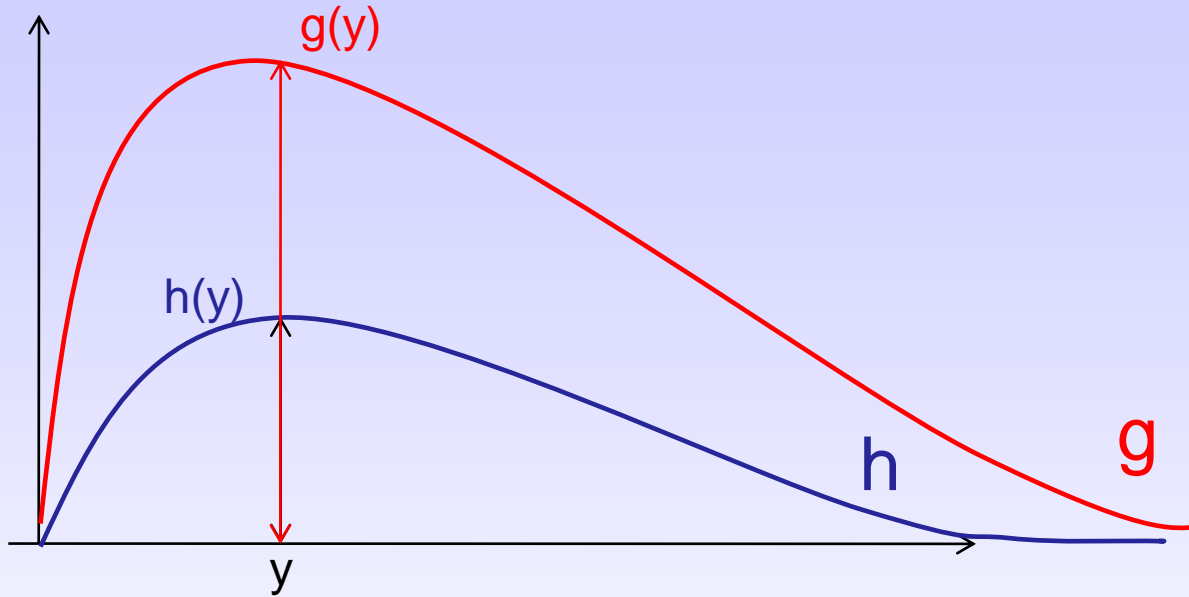
1-Tomamos otra densidad g (la envoltura) que sea fácil de simular:

$$\frac{g(y)}{\int_{y \in \text{Sop}(g)} g(y) dy}, \quad \text{Sop}(h) \subseteq \text{Sop}(g), g(x) \geq h(x)$$

2-Generamos un número $y \sim g$, y lo aceptamos con probabilidad $h(y)/g(y)$. En caso de rechazo repetimos el proceso.



5. Métodos estándar de generación de n° aleatorios. Método del rechazo.



Algoritmo para el método del rechazo:

- 1-Simular un valor de $y \sim g(y) / \int_{y \in \text{Sop}(g)} g(y) dy$ y de $R \sim U(0,1)$
- 2-Si $R < h(y)/g(y)$ se acepta y .
En caso contrario se vuelve a 1.



5. Métodos estándar de generación de n^0 aleatorios. Generación de números aleatorios para F.D. estándar

Distribución logarítmico-normal:

Si $X \sim N(\mu, \sigma^2)$, $Y = e^X > 0$, entonces Y es logarítmico-Normal. Su densidad, media y varianza son respectivamente: (ϕ es la densidad $N(0,1)$)

$$f_Y(y) = \frac{1}{\sigma y} \phi\left(\frac{\log(y) - \mu}{\sigma}\right), x > 0, \quad \mu_Y = e^{\mu + \sigma^2/2}, \quad \sigma_Y = \mu_Y \sqrt{e^{\sigma^2} - 1}$$

Se simulan valores de X y se calcula $Y = e^X$

Distribución normal bivariada:

$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$ La correlación entre X, Y es $\rho \in [-1, 1]$,

Y las distribuciones de $Y|X$ e $X|Y$ son respectivamente:

$$N\left(\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x - \mu_1), \sigma_2^2(1 - \rho^2)\right), N\left(\mu_1 + \rho \frac{\sigma_1}{\sigma_2}(x - \mu_2), \sigma_1^2(1 - \rho^2)\right)$$



5. Métodos estándar de generación de n^0 aleatorios. Generación de números aleatorios para F.D. estándar

Para generar un par (X, Y) tenemos que generar $Z_1, Z_2 \sim N(0, 1)$ y luego calcular:

$$x = \mu_1 + \sigma_1 Z_1, \quad y = \mu_2 + \rho \sigma_2 Z_1 + \sigma_2 \sqrt{(1 - \rho^2)} Z_2$$

Matricialmente:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \sigma_1 & 0 \\ \rho \sigma_2 & \sigma_2 \sqrt{1 - \rho^2} \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$$



5. Métodos estándar de generación de n^0 aleatorios. Ejercicios

Ejercicios:

11: Simulación por el método de la inversa

12: Simulación por el Método del rechazo

13, 14, 15: Aplicación a Seguros: simulación del daño total y del número de siniestros.

18, 19: Aplicación a Finanzas: simulación del modelo de la caminata log-normal y del modelo del árbol binomial.



6. Métodos Monte Carlo

¿Qué es Monte Carlo?

Los métodos Monte Carlo sirven para estimar integrales cuando las resoluciones simbólicas y numéricas no aportan soluciones. El fracaso de estos métodos es debido fundamentalmente a una alta dimensionalidad de la integral que hace imposible su aproximación numérica.

Los métodos Monte Carlo vienen de los años 40 del siglo XX. Usados entonces por los físicos e ingenieros, actualmente tienen un campo de aplicación muy extenso e importante en Finanzas y Seguros.

Dentro de este último contexto, se simulan números aleatorios que se utilizan para calcular valores o parámetros importantes de un modelo. Por ejemplo:
Medidas del riesgo, valores de opciones, etc...



6. Métodos Monte Carlo

¿Qué es Monte Carlo?

Los métodos Monte Carlo consisten en :

1. Representar la integral que se desea aproximar como una esperanza matemática.
2. Construir un estimador (por ejemplo la media aritmética).
3. Simular la distribución.
4. Estimar la integral=esperanza matemática con la ayuda de la muestra obtenida, aplicando el estimador



6. Métodos Monte Carlo

¿Qué es Monte Carlo?

A parte de los puntos 1,2,3,4 anteriores, Hay otros dos temas que son cruciales:

- ¿Cuan buena es la estimación obtenida?
 - Veremos que esto depende de la varianza del estimador empleado.
 - Para mejorar los métodos Monte Carlo existen las técnicas de reducción de la varianza (que no estudiaremos aquí).
- En general se busca un método que sea eficiente, en el sentido de que produzca una estimación aceptable con el menor número posible de simulaciones
 - De esa forma se pretende minimizar el tiempo de ejecución.



6. Métodos Monte Carlo

Un primer ejemplo

Vamos a ver un ejemplo en un caso en que sí podemos obtener una aproximación numérica del valor requerido. De esta forma podremos hacer comparaciones entre los dos valores:

$$I_{\alpha} = \int_0^{\infty} x^{\alpha-1} e^{-x} dx = \Gamma(\alpha), \alpha > 0.$$



funcion gamma

Trabajaremos con el siguiente valor de la función Γ :

$$\Gamma(1.9) = 0.96176$$



6. Métodos Monte Carlo

Un primer ejemplo

Aplicando la idea que acabamos de exponer (punto 1), podemos darnos cuenta de que tenemos una densidad exponencial:

$$f(x) = e^{-x}, x \geq 0 \Rightarrow I_\alpha = E_f(X^{\alpha-1})$$

Luego para estimar la integral podemos llevar a cabo el resto de los puntos 2,3,4, diseñando el experimento:

• Construcción del estadístico:
(es un estimador insesgado de I_α)

$$\hat{I}_\alpha = \frac{1}{n} \sum_{i=1}^n X_i^{\alpha-1}$$

• Simulación una muestra de números independientes de la densidad f : X_1, \dots, X_n

En cuanto a la varianza del estimador tenemos que:



6. Métodos Monte Carlo

Un primer ejemplo

- Varianza del estimador \hat{I}_α

$$\text{Var}_f(\hat{I}_\alpha) = \frac{\text{Var}_f(X^{\alpha-1})}{n} \Rightarrow \sigma_f(\hat{I}_\alpha) = \frac{\sigma_f(X^{\alpha-1})}{\sqrt{n}} \quad (\text{Error estandar e.e.})$$

- El e.e. varía con el tamaño muestral según un factor de $n^{-1/2}$:
POR TANTO:

Si queremos disminuir el e.e. en un factor de 10^{-k} , necesitamos aumentar el tamaño muestral en 10^{2k} .

- Ejemplo: $k=5$ entonces $2k=10$.

$10^{10} = 1.000.000.000$ de simulaciones.

- Conclusión:

La precisión extra puede implicar un coste (tiempo de ejecución) desproporcionado.



6.Métodos Monte Carlo

Un primer ejemplo

Aplicando el anterior plan de trabajo a nuestra integral I_α :

Ejemplo7: Estimación de Gamma mediante Monte Carlo.



Monte Carlo
Gamma



Ejercicios

Ejercicios:

18, 19: Integración Monte Carlo



Bibliografía

Adler, J. (2012): R in a nutshell. O'Reilly Media Inc.

[http://web.udl.es/Biomath/Bioestadistica/R/Manuals/r in a nutshell.pdf](http://web.udl.es/Biomath/Bioestadistica/R/Manuals/r%20in%20a%20nutshell.pdf)

Atkinson, K.E. (1989): An Introduction to numerical analysis. 2nd Ed. John Wiley.

Dagpunar, J. (2007): Simulation and Monte Carlo. With applications in finance and MCMC. John Wiley.

Henrici, P. (1962): Discrete Variable Methods in Ordinary Differential Equations . John Wiley.

L'écuyer P., Panneton F. (2007): F_2 -Linear Random Number Generators.

Matsumoto M., Nishimura T.(1997): Mersenne Twister: A 623- dimensionally equidistributed uniform pseudorandom number generator

Venables, Smith (2011): An introduction to R.

<http://cran.r-project.org/doc/manuals/R-intro.pdf>

The Mersenne Twister Home Page:

<http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/emt.html>

Informática y Cálculo Numérico. Ejercicios.

Prof. José Luis Vilar Zanón
Departamento Economía Financiera y Actuarial y Estadística
Máster en Ciencias Actuariales y Financieras
Universidad Complutense de Madrid

5 de septiembre de 2023

Resumen

Hoja de ejercicios de la asignatura. Las entregas deberán ser hechas en las fechas fijadas en el campus virtual.

Índice

1. Aproximación de raíces de ecuaciones.	2
2. Ecuaciones diferenciales de primer orden	3
3. Integración numérica.	4
4. Generadores de números aleatorios $U(0, 1)$. Simulación de números aleatorios.	4
5. Simulación de fenómenos aleatorios en seguros y finanzas	7
6. Integración Monte Carlo	10

1. Aproximación de raíces de ecuaciones.

Ejercicio 1. Método de Newton para aproximar una raíz.

Consideremos una ecuación $F(x) = 0$. Se trata de investigar que funciones R permiten la aproximación numérica de raíces de ecuaciones. Para ello se puede usar, por ejemplo, el paquete **rootSolve**.

Pruebe con las funciones de dicho paquete aproximando las raíces de las siguientes ecuaciones: (Si hace falta una semilla se puede por ejemplo, representar gráficamente las funciones para obtenerla de forma visual)

1. $F(x) = x^4 + 4x^3 - 7x^2 - 22x + 24$

2. $F(x) = 10 \log(x) - \sqrt{x}$

3. Aproxime la raíz del siguiente sistema tomando el punto $(1, 1, 1)$ como semilla:

$$\begin{cases} x_1 + x_2^2 + 2x_3 - 10 = 0 \\ x_1 - x_1x_2 + x_3 - 5 = 0 \\ 2x_1^2 - x_2^2 + x_3 - 10 = 0 \end{cases}$$

Ejercicio 2. Aplicación del método de Newton: Cálculo de la TIR-Tasa Interna de Retorno.

Considere el tiempo dividido en periodos $t = 0, 1, 2, \dots$, en donde el periodo $t = 0$ representa el presente. Un negocio, un programa de inversiones etc, puede ser representado mediante un flujo de caja que sitúa un capital x_i en cada uno de los periodos $t = 0, 1, \dots, n$:

$$(x_0, x_1, \dots, x_n)$$

Un capital negativo representará una salida de dinero de nuestro bolsillo (inversión, costes etc. . .) mientras que uno positivo representa un ingreso. Para valorar estos flujos de caja disponemos de dos técnicas sencillas: la del valor actual neto (VAN) y la de la tasa interna de retorno (TIR). Centrémonos en esta última.

Considerando el interés compuesto $\lambda > 0$ entre periodos, si actualizamos los capitales del flujo de caja al periodo presente $t = 0$ obtenemos la siguiente expresión:

$$\sum_{k=0}^n \frac{x_k}{(1 + \lambda)^k}. \quad (1)$$

Pues bien la TIR se define como el valor de $\lambda > 0$ que anula la anterior expresión:

$$\exists? \lambda > 0 / \sum_{k=0}^n \frac{x_k}{(1 + \lambda)^k} = 0. \quad (2)$$

Escriba un código R que calcule la TIR y aplicarlo a los siguientes flujos de caja:

1. $(-10, 20)$

2. $(-10, 12)$

3. $(-5, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10)$

4. $(-2, 0, 0, 5, 0, 0, 5, 0, 1, 2, 3, 4)$

Ejercicio 3. * Aplicación del método de Newton: Estimación máximo verosímil.

La aproximación de raíces también es un paso necesario en Estadística, por ejemplo, en la Estimación Puntual y más concretamente en el cálculo de los Estimadores Máximo Verosímiles (EMV) de los parámetros de algunas distribuciones.

Tomemos por ejemplo el fenómeno del número de siniestros producidos durante un periodo de tiempo en una cartera de pólizas de seguros. Llamemos n_k al número de pólizas que han sufrido $k = 0, 1, 2, \dots$ siniestros en el periodo (estas son las *frecuencias observadas*). Cuando hacemos la hipótesis de que la muestra ha sido producida por una distribución de una determinada familia paramétrica, podemos determinarla estimando puntualmente los parámetros de la familia a partir de la muestra observada. Veamos un ejemplo: el caso en que la familia paramétrica es la Binomial Negativa.

$$f(k|\alpha, \beta) = \frac{\Gamma(k + \alpha)}{\Gamma(\alpha)k!} \left(\frac{\beta}{1 + \beta}\right)^k \left(\frac{1}{1 + \beta}\right), \quad k = 0, 1, 2, \dots \quad (3)$$

A partir de una muestra (n_k) podemos calcular la media y la varianza muestrales \bar{n}, S^2 . Entonces los EMV's $\hat{\alpha}, \hat{\beta}$ son las raíces del siguiente sistema de ecuaciones, siendo $n = \sum n_k$ el número total de siniestros de la cartera durante el periodo de tiempo:

$$\hat{\alpha} > 0 / n \log \left(1 + \frac{\bar{n}}{\hat{\alpha}}\right) - \sum_{k=1}^{\infty} n_k \sum_{m=0}^{k-1} \frac{1}{\hat{\alpha} + m} = 0 \quad (4)$$

$$\hat{\beta} = \frac{\bar{n}}{\hat{\alpha}} \quad (5)$$

1. Construya una función R para la estimación máximo verosímil y aplíquela a las siguientes muestras:

a) $(19, 798; 4, 141; 845; 163; 44; 9)$, $k = 0, \dots, 5$.

b) $(8, 283; 1, 138; 347; 143; 53; 23; 8; 3; 1; 0; 1)$, $k = 0, \dots, 10$.

Observaciones 1. Si la función R para aproximar la raíz de la primera ecuación de (4) pide una semilla, se puede dar la *estimación por momentos* del parámetro α : $\bar{\alpha} = \frac{\bar{n}^2}{S^2 - \bar{n}}$.

2. Ecuaciones diferenciales de primer orden

Ejercicio 4. Poligonal de Euler: Aproximación numérica de la solución de un PVI-Problema de Valor Inicial.

Para los siguientes problemas de valor inicial (PVI), calcule las soluciones aproximadas sobre un intervalo con extremo izquierdo en el 0, correspondientes a los pasos $dt = 1; 0.5; 0.1; 0.01$, y represente dichas aproximaciones gráficamente:

1. $dy = 0.3ydt$, $y(0) = 10$

2. $dy = (0.3y^2 - 4t) dt$, $y(0) = 1$

3. $dy = (y^2 + t^2) dt$, $y(0) = -1$

4. $dy = (1 + 2ty) dt, \quad y(0) = 1$

Observaciones 2. Las Ecuaciones Diferenciales Ordinarias 2, 3 no pueden ser resueltas simbólicamente. La ecuación 1 es la del crecimiento exponencial y la solución de la 4 es

$$y(t) = e^{-t^2} \int_0^t e^{s^2} ds,$$

cuya primitiva no se puede expresar en términos simbólicos más allá de la propia integral que la define. Se trata por tanto de una expresión para la que es necesaria aplicar Cálculo Numérico si queremos calcular sus valores (integración numérica en este caso).

3. Integración numérica.

Ejercicio 5. Integración numérica en R.

Calcule numéricamente las siguientes integrales con ayuda de la función `integrate()`:

1. Para una variable aleatoria $X \sim N(0, 1)$: $P\{X \geq 4\}$
2. Gamma truncada en 4: $\Gamma(\alpha, a) = \int_a^\infty x^{\alpha-1} e^{-x} dx, \quad a = 4, \quad \alpha = 3.$
3. $\int_{-\infty}^\infty e^{-x^2} |\cos(x)| dx$
4. $\int_{-\infty}^\infty e^{-|x|} e^{-\frac{x^2}{2}} dx$

Observaciones 3. En los ejercicios 18 y 19 aproximaremos las dos últimas integrales mediante el método Monte Carlo, y compararemos los resultados. Así mismo, en las diapositivas de la asignatura usaremos el caso 2 como un caso sencillo con el cual ejemplificar la integración Monte Carlo.

4. Generadores de números aleatorios $U(0, 1)$. Simulación de números aleatorios.

Ejercicio 6. GLC -Generadores Lineales Congruentes.

Genere mediante código R el ciclo completo para los generadores lineales congruentes siguientes. Tome nota del periodo y compruebe que está de acuerdo con la teoría.

1. $X_i = (5X_{i-1} + 3) \text{ mód } 2^4 \in [0, 2^4 - 1], \quad i = 1, 2, \dots, \quad X_0 = 5$
2. $X_i = (5X_{i-1} + 3) \text{ mód } 2^4 \in [0, 2^4 - 1], \quad i = 1, 2, \dots, \quad X_0 = 7$
3. $X_i = (7X_{i-1} + 3) \text{ mód } 2^4 \in [0, 2^4 - 1], \quad i = 1, 2, \dots, \quad X_0 = 5$
4. $X_i = (5X_{i-1} + 4) \text{ mód } 2^4 \in [0, 2^4 - 1], \quad i = 1, 2, \dots, \quad X_0 = 5$
5. $X_i = (5X_{i-1}) \text{ mód } 64 \in [0, 63], \quad i = 1, 2, \dots, \quad X_0 = 3$
6. $X_i = (5X_{i-1}) \text{ mód } 64 \in [0, 63], \quad i = 1, 2, \dots, \quad X_0 = 4$

Ejercicio 7. (¿Consiguen engañarnos los generadores lineales congruentes? (1). Considere el generador lineal congruente de periodo máximo

$$X_i = (aX_{i-1} + c) \pmod{m} \quad (6)$$

1. Demuestre que tomando el ciclo completo de números generados por (6), $(X_i)_{i=1}^m$, se tiene lo siguiente:

$$E(X_i) = \frac{m-1}{2}, \quad \text{Var}(X_i) = \frac{(m-1)^2}{12}$$

Para ello utilice los siguiente resultados conocidos:

$$\sum_{i=0}^{k-1} i = \frac{k(k-1)}{2}, \quad \sum_{i=0}^{k-1} i^2 = \frac{k(k-1)(2k-1)}{6}$$

A continuación escriba $R_i = \frac{X_i}{m}$, y pruebe que

$$\lim_{m \rightarrow \infty} E(R_i) = \frac{1}{2}, \quad \lim_{m \rightarrow \infty} \text{Var}(R_i) = \frac{1}{12}$$

2. Busque y escriba la media y la varianza de una distribución uniforme (0,1). Compárelas con los resultados obtenidos en el apartado anterior. ¿Qué conclusión se puede sacar?

Observaciones 4. Este ejercicio sirve para comprobar que desde el punto de vista de las dos primeras características muestrales (media y varianza), los GCL suministran una secuencia de números que, cuanto mayor sea el periodo m , más se comporta como una muestra de una distribución Uniforme (0,1). Luego desde este punto de vista media-varianza bastaría con tomar un GCL de periodo muy grande para poder “creer” que su secuencia es una muestra de la anterior distribución.

Ejercicio 8. ¿Consiguen engañarnos los generadores lineales congruentes? (y 2).

Un posible test para saber si un generador de números aleatorios genera una secuencia que efectivamente puede pasar como una muestra de una distribución Uniforme (0,1) consiste en tomar la secuencia y someterla a una prueba de bondad del ajuste.

1. El profesor se ha inventado una secuencia de 10,000 números supuestamente distribuidos $U(0,1)$ e independientes (que omitimos por brevedad), y una vez reagrupados los valores en subintervalos, se han producido las siguientes frecuencias:

Intervalo	0-0.1	0.1-0.2	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1
Frecuencia	1023	1104	994	993	1072	930	1104	969	961	850

- a) Aplique una prueba de bondad de la χ^2 para la hipótesis nula de que esta muestra procede de una distribución $U(0,1)$. Compruebe entonces que la hipótesis nula puede ser rechazada.
2. Para comprobar de qué forma un GCL pasa este test de uniformidad:

- a) Genere 10,000 números aleatorios en el intervalo $(0, 1)$, utilizando el GCL del ejemplo 2 de la página 44 de la Teoría, y el código R para GCLs. Clasifíquelos por intervalos igual que en la tabla anterior y calcule las frecuencias de cada intervalo.
- b) Aplique una prueba de bondad de la χ^2 para la hipótesis nula de que esta muestra procede de una distribución $U(0, 1)$. Compruebe que la hipótesis nula no puede ser rechazada.

Observaciones 5. Este ejercicio sirve para comprobar que los GCL también aprueban el test de la prueba de bondad del ajuste de la χ^2 , siempre que tengan un periodo suficientemente grande.

Ejercicio 9. ¿Consiguen engañarnos los generadores lineales congruentes? (y 3).
Obtenga las secuencias completas de números correspondientes a los generadores

$$X_i = (7X_{i-1}) \pmod{13}, \quad Y_i = (Y_{i-1} + 5) \pmod{16}$$

Representando los pares respectivos $\{(X_i, X_{i+1})\}$, $\{(Y_i, Y_{i+1})\}$ $i = 1, 2, \dots$, podríamos esperar, en el caso en que ambos GCL suministrasen efectivamente secuencias aleatorias $U(0, 1)$, que ambos dibujos fuesen nubes de puntos sin ninguna organización dentro del cuadrado $(0, 1)^2$.

Represente las dos nubes, y en función del dibujo obtenido discuta la aleatoriedad de ambas secuencias. ¿Cuáles son los periodos de estos generadores? ¿Qué le sucedería a un GCL de periodo más largo tal como el del ejemplo 2 página 44 de la Teoría?

Observaciones 6. En este ejercicio comprobamos que los GCL no superan el llamado *test de uniformidad* en dimensión 2. Desde la década de los noventa del siglo XX se han conseguido definir otros generadores de números aleatorios que superan este test en muy alta dimensión. Tal es el caso del generador *Mersenne-Twister*, presentado en la Teoría páginas 48-51, que pasa el test hasta la dimensión 623. Por este motivo este es el generador de números aleatorios de uso corriente hoy en día.

Ejercicio 10. Simulación de modelos de probabilidad con R.

Con ayuda de las funciones R, simular 10^6 números aleatorios pseudo distribuidos según las distribuciones que se indican abajo, correspondientes a la semilla 567.

Una vez simulada cada muestra efectuar las siguientes tareas:

1. Ordenar la muestra.
2. Recuperar el número que ocupa la posición 112,345.
3. Recuperar los números $0,9 < z < 1$.
4. Recuperar los números $z > 3$.
5. Calcular los cuantiles de la muestra 99 %, 95 %, 1 %

Las distribuciones son las siguientes:

1. $LN(\mu = 2, \sigma = 1/2)$.
2. χ^2 con 2 grados de libertad.
3. Inversa Gaussiana de media 2 y desviación típica 1/2.

4. Poisson con $\lambda = 2$.
5. BN($r = 1.5, \beta = 1.7$).

Ejercicio 11. Simulación por el método de la inversa.

Usar la inversa de la función de distribución para desarrollar métodos de simulación para las siguientes funciones de densidad:

1. $\frac{1}{20}$, $x \in [5, 25]$ (Uniforme).
2. $\frac{x-2}{8}$, $x \in (2, 6]$
3. $\frac{1}{5}e^{-\frac{x}{5}}$ $x \in (0, \infty]$ (Exponencial con media 5)

Ejercicio 12. Simulación por el método del rechazo.

1. Utilice el algoritmo del rechazo para producir números aleatorios de una densidad proporcional a $h(x)$. Use la envolvente $y = g(x)$ indicada.
 $h(x) = x(1-x)$ sobre el intervalo $(0, 1)$, con la envolvente $g(x) = cte \in \mathbb{R}$. (Tendrá que determinar el mejor valor de esa “cte” para llevar a cabo la simulación)
2. Realice 100 simulaciones y anote el número de valores aceptados y rechazados.
3. Represente gráficamente las funciones $h(x), g(x)$ y los puntos simulados sobre el intervalo $(0,1)$ contenido en el eje de abscisas.
4. Basándose en los resultados obtenidos en los apartados 2 y 3, concluya acerca de la calidad de este generador de números aleatorios.

Observaciones 7. Aunque es claro que el modelo h es una Beta, no se trata en este ejercicio de utilizar el generador de la distribución “Beta” de R. Siga los pasos indicados en los apartados aplicando el algoritmo explicado en la página 55 de la teoría.

5. Simulación de fenómenos aleatorios en seguros y finanzas

Ejercicio 13. Aplicación a los seguros: simulación del daño total de una cartera de pólizas.

El *daño total o siniestralidad total* provocada por una masa de pólizas de seguros (ya sea esta una sola póliza o una cartera de estas) durante el próximo periodo es una variable aleatoria (v.a.) S compuesta:

$$S = \sum_{i=1}^N X_i, \quad (S = 0, \text{ si } N = 0), \quad (7)$$

siendo:

- N v.a. del *número de siniestros* durante el periodo, o v.a. *primaria*.
- $(X_i)_{i=1}^{\infty}$ variables aleatorias independientes idénticamente distribuidas (v.i.i.d) de las *cuantías individuales de los siniestros*, o v.a. *secundaria*.
- N, X_i son v.a. independientes.

Se trata pues de un modelo probabilístico para las *sumas aleatorias de variables aleatorias*, el más importante de los modelos de riesgo en seguros. Supongamos a continuación que $N \sim P(10)$, $X \sim \text{Gamma}(\alpha = 3.2, \beta = 0.4)$.

1. Simule una muestra de 100,000 daños totales:
 - a) Escriba primero el algoritmo.
 - b) Codifique el algoritmo.
2. Represente gráficamente la función de distribución muestral del daño total en un periodo de tiempo.
3. ¿Cuál es el daño total máximo?

Observaciones 8. Para la simulación utilícese la semilla 1000.

Ejercicio 14. * Simulación del número de siniestros de una cartera de pólizas con el modelo del riesgo bueno-riesgo malo.

Un recurso muy importante para modelizar la *heterogeneidad*¹ de una cartera de pólizas de seguros, base para la *tarificación a posteriori* en seguros generales, es el de la distribución de *Poisson ponderada*, que se utiliza para modelizar el número de siniestros $N \in \{1, 2, 3, \dots\}$ producidos en la cartera durante un periodo de tiempo. Su función de cuantía es la siguiente:

$$p_n = P\{N = n\} = \int_0^{+\infty} \frac{\lambda^n}{n!} e^{-\lambda} dU(\lambda), \quad n = 0, 1, 2, \dots \quad (8)$$

La v.a. Λ representa el número medio de siniestros por unidad de tiempo (por ejemplo, un año), y cada póliza de la cartera puede tener un valor particular $\Lambda = \lambda$ con una probabilidad $dU(\lambda)$, lo que da como resultado una cartera heterogénea en la que cada póliza tiene su propio número medio de siniestros/año λ .

Cuando la v.a. Λ es discreta, si llamamos $dU(\lambda) = u(\lambda)$ a su función de cuantía y $\text{sop}(U) = \{\lambda > 0 / u(\lambda) > 0\}$ al conjunto de los valores de con probabilidad distinta de cero, la expresión (1.9) se reduce a:

$$p_n = P\{N = n\} = \sum_{\lambda \in \text{sop}(U)} \frac{\lambda^n}{n!} e^{-\lambda} u(\lambda), \quad n = 0, 1, 2, \dots \quad (9)$$

Tomemos una cartera compuesta en un 90 % de pólizas “buenas” con una media de siniestralidad de $\lambda = 0.08$ siniestros por año, y en otro 10 % de pólizas “malas” con una media de siniestralidad de $\lambda = 0.3$ siniestros al año). Esto también puede expresarse como

$$u(\lambda) = \begin{cases} 0.9, & \lambda = 0,08 \\ 0.1, & \lambda = 0,3 \end{cases}$$

Supongamos además que dicha cartera está compuesta por 50.000 pólizas.

1. Simule el número de siniestros de esta cartera: escriba el algoritmo y luego codifíquelo.

¹Esto consiste en que las pólizas de la cartera no son homogéneas en el riesgo que representan para la compañía de seguros. Se modeliza esta característica suponiendo que cada póliza tiene un número medio de siniestros anual distinto y aleatorio por ser este desconocido a priori

2. Agrupe el resultado obtenido en un vector de *frecuencias observadas*.

Observaciones 9. El apartado 2 es para darle a la muestra obtenida la forma $(n_0, n_1, \dots, n_k, \dots)$ siendo n_k el número de pólizas de la cartera que han sufrido k siniestros durante el año. Para la simulación utilice la semilla 789.

Ejercicio 15. * Simulación del número de siniestros de una cartera de pólizas con el modelo de la binomial negativa.

Continuando con el caso general de las distribuciones de Poisson ponderadas (8), supongamos ahora que la v.a. Λ es absolutamente continua. Si llamamos $u(\lambda)$ a su función de densidad, la expresión (8) es entonces:

$$p_n = P\{N = n\} = \int_0^{+\infty} \frac{\lambda^n}{n!} e^{-\lambda} u(\lambda) d\lambda, \quad n = 0, 1, 2, \dots \quad (10)$$

Tomemos una cartera para la que la v.a. $\Lambda \sim \text{gamma}(0.234, 0.345)$. Supongamos además que dicha cartera está compuesta por 50.000 pólizas.

1. Simule el número de siniestros de esta cartera: escriba el algoritmo y luego codifíquelo. Utilice la semilla 999.

Observaciones 10. Ponga cuidado de que la muestra obtenida tenga la forma $(n_0, n_1, \dots, n_k, \dots)$ siendo n_k el número de pólizas de la cartera que han sufrido k siniestros durante el año.

Ejercicio 16. * Aplicación a finanzas: el modelo de la caminata logarítmico-normal.

El modelo de Black-Scholes para simular el precio de una acción $X = \{X(t) : t \geq 0\}$ viene dado por:

$$X(t) = X(0) \exp \left(\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma \sqrt{t} Z \right), \quad (11)$$

siendo $X(0)$ el precio actual, μ la tendencia, σ la volatilidad y $Z \sim N(0, 1)$.

1. Utilice las expresiones de la media y la varianza de una distribución logarítmico normal para probar que la media y la desviación típica de $X(t)$ vienen dados por:

$$E(X(t)) = X(0)e^{\mu t} \quad (12)$$

$$\sigma(X(t)) = X(0)e^{\mu t} \sqrt{e^{\sigma^2 t} - 1} \quad (13)$$

2. Escriba un código R que simule 1000 valores de $X(0.5)$ cuando $X(0) = 100\text{€}$, $\mu = 0.1$, $\sigma = 0.3$.

3. Suponiendo que en $t = 0.5$ el valor liquidativo (o liquidación) de la opción Put con strike (precio de ejercicio) $K = 100$ es

$$V(X(0.5)) = \text{máx}\{100 - X(0.5), 0\},$$

estime la liquidación media.

4. Suponga que el precio de dos acciones A y B vienen dados por:

$$X_A(t) = 10 \exp \left(\left(0.1 - \frac{0.09}{2} \right) t + 0.3 \sqrt{t} Z_A \right), \quad \mu = 0.1, \quad \sigma = 0.3 \quad (14)$$

$$X_B(t) = 10 \exp \left(\left(0.08 - \frac{0.09}{2} \right) t + 0.3 \sqrt{t} Z_B \right), \quad \mu = 0.08, \quad \sigma = 0.3 \quad (15)$$

siendo $Z_A, Z_B \sim N(0, 1)$, $\rho = 0.8$. suponga también que la liquidación de una opción es:

$$V(X(0.5)) = \text{máx}\{50 - X_A(0.5) - 2X_B(0.5), 0\}$$

Estime el valor liquidativo medio de la opción usando valores simulados de dicho valor liquidativo.

Observaciones 11. Para generar las simulaciones de pares normales (Z_A, Z_B) correlados $\rho = 0,8$, utilice el método para la distribución normal bivariada explicado en las páginas 56-57 de la teoría.

Ejercicio 17. *Aplicación a finanzas: el modelo del árbol binomial.

El modelo del árbol binomial es un modelo en tiempo discreto usado comúnmente para reproducir la evolución de los precios de los activos que se negocian en los mercados, tales como las acciones. Llamemos S_i al precio del activo en el periodo ih , $i = 0, 1, 2, \dots$, siendo $h > 0$ un incremento en el tiempo. Sean μ, σ la tasa de crecimiento medio y la volatilidad respectivamente. Definimos:

$$u = \frac{1}{2} \left(e^{-\mu h} + e^{(\mu + \sigma^2)h} \right) + \frac{1}{2} \sqrt{(e^{-\mu h} + e^{(\mu + \sigma^2)h})^2 - 4} \quad (16)$$

$$d = u^{-1} \quad (17)$$

$$p = \frac{e^{\mu h} - d}{u - d} \quad (18)$$

Entonces:

$$S_i = X_i S_{i-1}$$

Siendo X_i , $i = 0, 1, 2, \dots$ unas v.a.s Bernoulli independientes con distribución:

$$P(X_i = u) = p, \quad P(X_i = d) = 1 - p, \quad \forall i$$

1. a) Simule los precios al final de cada semana durante un año. Tome $S_0 = 100$, $\mu = 0,2$ (anual), $\sigma = 0,3$ (anual), $h = 1/52$.

Observaciones 12. Presente los resultados en un array cuya primera fila sea las probabilidades y la segunda los precios.

6. Integración Monte Carlo

Ejercicio 18. *Aproximación de una integral por el método Monte Carlo.

Utilice un método Monte Carlo basado en 1000 simulaciones de una $N(\mu, \sigma)$ para encontrar un intervalo de confianza al 95% para la integral

$$\int_{-\infty}^{\infty} e^{-x^2} |\cos(x)| dx,$$

que ya vimos en el ejercicio 5.

Ejercicio 19. *Aproximación de una integral por el método Monte Carlo.

Volvamos a otra de las integrales planteadas en el ejercicio 5:

$$\int_{-\infty}^{\infty} e^{-|x|} e^{-\frac{x^2}{2}} dx$$

1. Plantee su estimación mediante el método Monte Carlo.
2. Generando una muestra de 10 simulaciones, estime θ . (Utilice la semilla 123 para generar los números aleatorios)
3. ¿Cuántas simulaciones serían necesarias para reducir el error estándar a 0.001?
4. Estime de nuevo la integral con el número de simulaciones obtenido en el apartado anterior. (Utilice de nuevo la semilla 123)