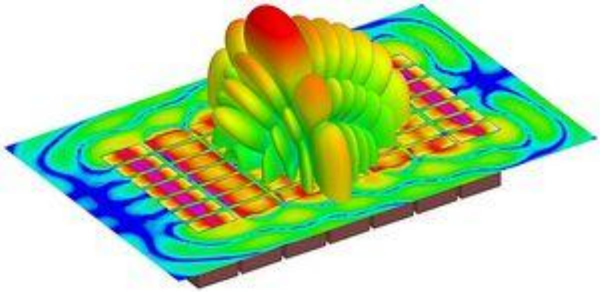


Simulación

Simulación

- Corresponde a la representación abstracta de un **sistema** en un computador.
- Tiene por objetivo emular el funcionamiento del sistema.
- Se usa para evaluar numéricamente el sistema bajo ciertas condiciones.
- ¿Por qué realizarla?
  - Permite evaluar muchas alternativas de forma rápida.
  - **COSTO.**
  - Facilidad de comprensión.

Sistemas

- Un sistema corresponde a un conjunto de entidades (máquinas, personas, átomos, etc.) que interactúan para lograr algún fin lógico.
  - Ej: En una cola de un banco, el sistema son los clientes, los cajeros, las características de la cola, etc.

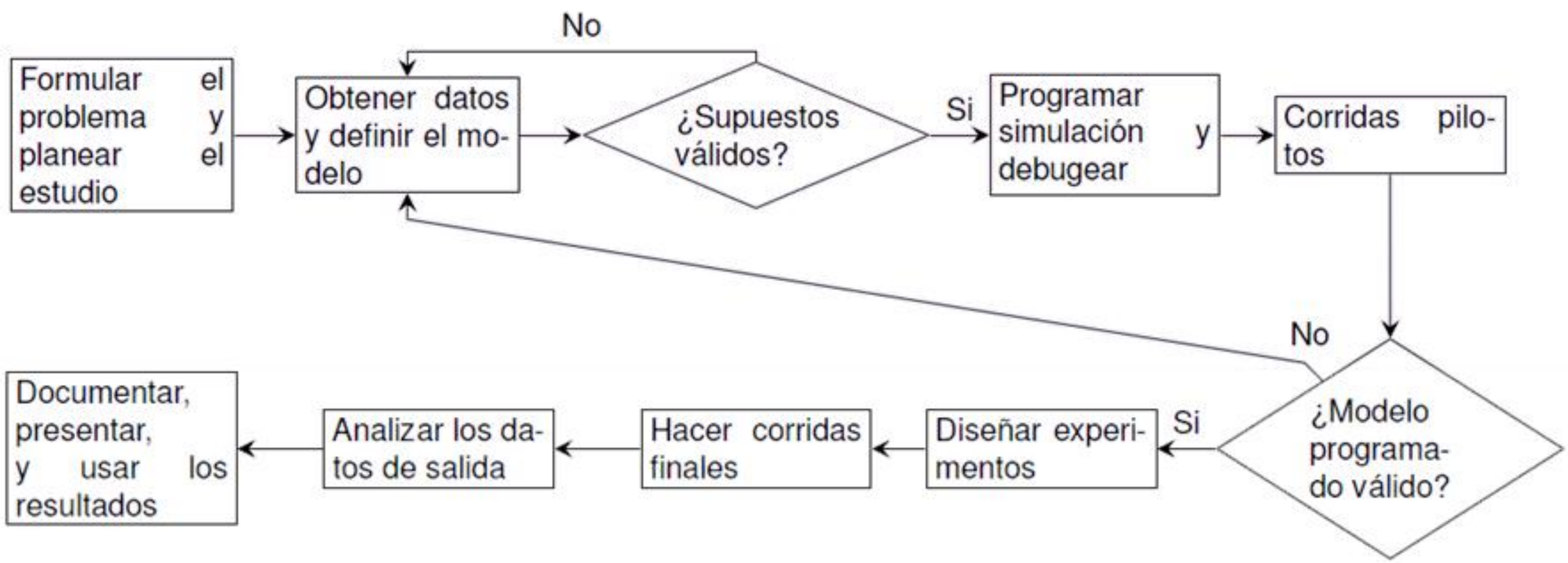
- Se caracteriza a través de estados, es decir, un conjunto de variables que describen un sistema en algún momento particular.
  - En el banco serían la cantidad de clientes en cola, el número de cajeros, y el estado de cada cajero.

Caracterización de Sistemas

- Los sistemas pueden ser dinámicos o estáticos.
  - Dinámicos: el sistema evoluciona a medida que pasa el tiempo.
    - Ej: Las colas del banco.
  - Estáticos: el tiempo no juega ningún factor.
    - Ej: Estimar utilidades del ejercicio. Simulaciones tipo Montecarlo en general.

- Los sistemas pueden ser discretos o continuos.
  - Discretos: Si las variables de estado cambian acorde a un conjunto de instantes numerables.
    - Ej: Ventas diarias en un supermercado.
  - Continuos: Si las variables de estado cambian en cada instante.
    - Ej: Posición de los planetas en el sistema solar.

Qué Necesitamos para Simular



Generación de Números Aleatorios

Generando Números Aleatorios

- El primer paso para poder obtener valores consistentes es la generación de números aleatorios.

- Partiremos asumiendo que existe un generador de números uniforme $g(\cdot)$ entre 0 y 1.
  - Este es un problema muy complejo.
    - ¿Qué es realmente aleatorio?
  - Los PC usan el reloj (al milisegundo) del computador.
  - Fourmilab en Suiza ofrece servicio basado en el decaimiento radioactivo de partículas: HotBits (http://www.fourmilab.ch/hotbits/)
  - Random.org usa el ruido atmosférico que captura en su sede.

Generación de Números Aleatorios (II)

- Lo que buscaremos es construir un «generador de números aleatorios»

- Este será un programa que produzca una secuencia de números que siga una distribución específica y que posea la apariencia de aleatoriedad.

- Por ello a estos números generados se le llaman «pseudo-aleatorios».

Método de la Transformada Inversa

- Probablemente el método más usado es el de la transformada inversa.

- Se desean generar números que sigan una distribución $F(x) = P(X \leq x)$.
  - $F(x)$ debe ser estrictamente creciente, pero no necesariamente continua (puede ser de tipo escalón).

- Para generar números aleatorios se sigue el siguiente procedimiento:
  1. Generar $u = g(\cdot)$.
  2. Sea $r = \inf\{x \mid u \geq F(x)\}$, x será un número pseudo-aleatorio obtenido a partir de la distribución $F(x)$.
     - En el caso continuo, $r = F^{-1}(u)$.

Método de Aceptación / Rechazo

- El método de la transformada inversa posee un problema: algunas veces la **función de distribución acumulada no es invertible**.

- En estos casos se aplica el método aceptación rechazo. Se considerar f : $S \to [0,1]$ función de densidad.
  1. Generar $c = \max\{f(x): x \in S\}$.
  2. Generar $x = g(\cdot)$ y ajustar x a conjunto S.
  3. Generar $Y = g(\cdot)$ y ajustar a intervalo $[0, c]$.
  4. Si $Y \leq f(x)$, retornar x.
  5. Si no, volver a 1.

Simulación de Montecarlo

Simulación de Montecarlo

- Corresponde al método de simulación estática más conocido y utilizado.
  - Usos van desde el riesgo de un proyecto hasta física de partículas.
- Parte de los siguientes supuestos:
  - Se tiene un fenómeno que resulta de una función de variables aleatorias independientes.
  - Si n es grande, se cumple que:

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < \frac{3\sigma}{\sqrt{n}}\right) \approx 99.8\%$$

  - Lo anterior sigue del teorema central del límite.

Simulación de Montecarlo (II)

- Podemos ahora definir un procedimiento para estimar el valor más probable de esta variable.
- Si $Y = f(X_1, X_2, \dots, X_n)$, con X_i variables aleatorias distribuidas según una ley F_{X_i}, en el rango $[l_i, u_i]$ el valor del fenómeno Y puede ser estimado como:

  1. Generar n vectores $(X_{1i}, X_{2i}, \dots, X_{ni})$ siguiendo la ley de probabilidad F_{X_i}, o en caso de no conocerla, generando números uniformes en cada intervalo $[l_i, u_i]$.
     - Lo anterior tiene sentido si n es grande, pues los números serán finalmente normales.
  2. Calcular $Y_i = f(X_{1i}, X_{2i}, \dots, X_{ni})$.
  3. El estimador más probable para Y será $\overline{Y(n)} = \sum_{i=1}^{n} \frac{Y_i}{n}$

Eligiendo el Número de Experimentos

- ¿Cuán grande debe ser n?
- Se puede determinar a partir de la expresión del error. De los supuestos:

$$\varepsilon = \frac{3\sigma}{\sqrt{n}} \Rightarrow n = \left(\frac{3\sigma}{\varepsilon}\right)^2$$

- Entonces, es posible estimar un límite para este error siguiendo estos pasos:
  1. Calcular $Y_{min} = \min\{f(X_1, \dots, X_n)\}$ y $Y_{max} = \max\{f(X_1, \dots, X_n)\}$. Esto se realiza a partir de los rangos de las variables.
  2. Calcular la desviación de estos dos valores y utilizar como sigma.
  3. Usar el promedio de estos dos valores y dividir por la tolerancia deseada.
  4. Usar fórmula para obtener un estimador de n.
  5. Una vez hechos los experimentos, calcular el error real a partir de la fórmula.

Ejemplo

- Se tiene un proyecto que utiliza 6 inputs, de distribución desconocida, tal que:

Activity	Minimum	Maximum
A	10,000	20,000
B	15,000	15,000
C	7,500	12,000
D	4,800	6,200
E	20,000	25,000
F	5,000	7,000
Total	62,300	85,200

- El valor del proyecto es la suma de estos seis valores.
  - $Y_{min} = 62{,}300, Y_{max} = 85{,}200$

Ejemplo (II)

- Para estimar el número de experimentos con un error de 2%.
  - $\sigma = 9.349$
  - Media $= 73.750$
  - $\varepsilon_{2\%} = \frac{73.750}{50} = 1.475$
  - $n = \left(\frac{3*9349}{1475}\right)^2 \approx 362.$
- Se realizan 362 corridas del experimento. En Excel.

Experimentos	362
Media	$ 73.993
Desv. Est.	$ 3.581
Error	$ 565
Error %	0,76%

Análisis de Resultados

- En general, se usa la media y la varianza.
  - El teorema central del límite nos asegura que estos convergerán a valores normales.

- Estimadores insesgados:
  - $\overline{X(n)} = \sum_{i=1}^{n} \frac{X_i}{n}$ es estimador insesgado de $E[X]$.
  - $S^2(n) = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X(n)}\right)^2}{n-1}$ es estimador insesgado de $Var(X)$ y se cumple que $Var\left(\overline{X(n)}\right) = \frac{Var(X)}{n}$.

Construcción de Intervalos de Confianza

- A partir de la media y la varianza encontrada podemos generar intervalos de confianza para nuestros resultados.
- Si hicimos suficientes experimentos (n suficientemente grande) se cumple que:

$$t_n = \frac{\overline{X(n)} - \mu}{\sqrt{\frac{S^2(n)}{n}}} \sim N(0,1)$$

  - Lo anterior es asintóticamente, pero se usa de forma general.
- Siguiendo eso, se tiene que:

$$P\left(-z_{1-\frac{\alpha}{2}} \leq t_n \leq z_{1-\frac{\alpha}{2}}\right) \approx 1 - \alpha$$

Construcción de Intervalos de Confianza (II)

- Con esto se construye el intervalo de confianza para μ (la media de nuestro experimento) como:

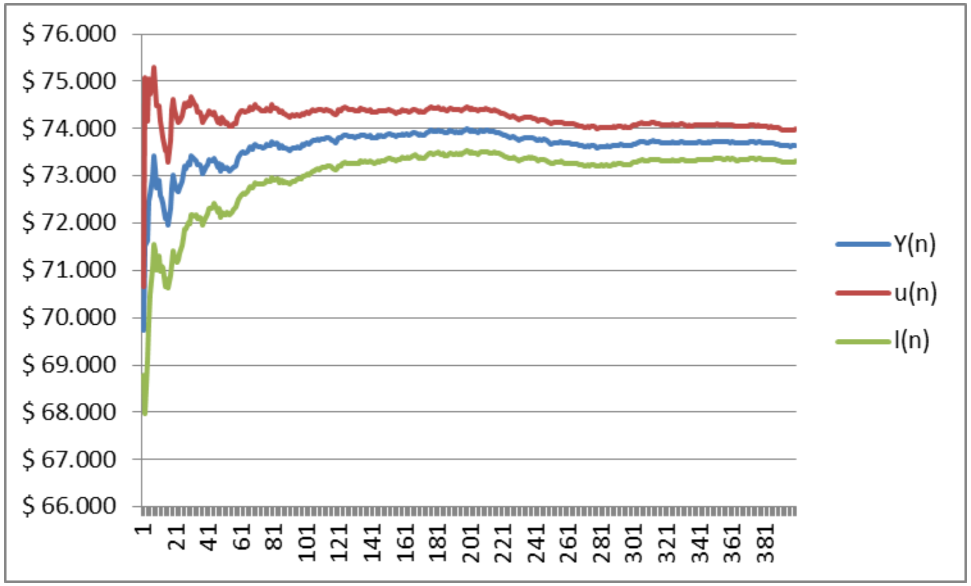
$$P\big(l(n) \leq \mu \leq u(n)\big) = 1 - \alpha$$

- Donde:
  - $l(n) = \overline{X(n)} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{S^2(n)}{n}}$
  - $u(n) = \overline{X(n)} + z_{1-\alpha/2}\sqrt{\frac{S^2(n)}{n}}$

- Lo anterior indica cuál es el área más probable (acorde a $1 - \alpha$) donde se encontrará el valor que buscamos.

En el Ejemplo

- Calculamos incrementalmente los valores de $\overline{Y(n)}, l(n), u(n)$.
- Graficamos estos valores para obtener los estimadores que se van obteniendo:



Métodos de Reducción de Varianza

Reducción de Varianza

- ¿Qué pasa si tenemos más información sobre una variable aleatoria?
- Pensemos en un evento Y de tal manera que lo podemos explicar como la resta de dos variables, X_1 y X_2.

$$Y = X_1 - X_2$$

- Se desea calcular $E[Y]$.
- Opción 1: Simular X_1 y X_2, y calcular $\bar{Y} = \bar{X}_1 - \overline{X_2}$.
  - En este caso se tiene que la varianza del estimador es $Var(\overline{X_1}) + Var(\overline{X_2})$.
- Opción 2: Simular $Z = X_1 - X_2$ directamente y calcular $\bar{Z}$.
  - La varianza es $Var(\bar{Z}) = \frac{1}{n}\left(Var(X_1) + Var(X_2) - 2Cov(X_1, X_2)\right)$
  - ¡Es menor mientras más grande sea $Cov(X_1, X_2)$!

Método de Variable de Control

- Este método se basa en agregar la información de una variable conocida.
- Si queremos estimar X, tal que existe una variable Z con dos características:
  - $E[Z]$ es conocida.
  - $Cov(X,Z) \neq 0$, es decir, existe una correlación entre ambas variables.
- Para aprovechar las características de Z simularemos la siguiente variable:

$$X_c = X + c(\bar{Z} - E[Z])$$

  - Este estimador es insesgado para todo valor de c.

Método de Variable de Control (II)

- La varianza de X_c corresponde a:
$$Var(X_c) = Var(X) + c^2 Var(Z) + 2c Cov(X, Z)$$
- ¿Qué valor debería tener c?
- Derivando e igualando a cero llegamos a:
$$c^* = -\frac{Cov(X, Z)}{Var(Z)}$$
- La varianza del estimador es:
$$Var(X_c) = Var(X) - \frac{Cov^2(X, Z)}{Var(Z)}$$
- Para aplicar el método se necesita estimar $Cov^2(X, Z)$ y $Var(Z)$.

Método de Variable de Control (III)

- Entonces, el procedimiento para estimar estos valores corresponderá a:
  1. Simular X y Z en p corridas piloto.
  2. Estimar $Cov(X,Z)$ como
  $$Cov(X,Z) = \frac{1}{p-1}\sum_{j=1}^{p}\left(X_j - \overline{X(p)}\right)\left(Z_j - \overline{Z(p)}\right)$$
  3. Estimar $Var(Z)$ como la varianza usual.
  4. Simular ahora n veces la variable
  $$X_c = X - \frac{Cov(X,Z)}{Var(Z)}(\bar{Z} - E[Z])$$
- Este estimador será insesgado y de mínima varianza.