Resumen 2do Parcial

Generación de números pseudoaleatorios

Un NÚMERO ALEATORIO es un número obtenido al azar.

Características:

- Que sean uniformemente distribuidos
- Estadísticamente independientes
- Reproducibles
- Periodo Largo
- Generados de manera rápida
- Que no requiera gran capacidad de almacenamiento

Métodos para encontrar números aleatorios:

- Manuales: moneda, dado
- De computación analógica: procesos físicos (corriente)
- De computación digital: mediante recurrencia
- Tablas de bibliotecas

Los números aleatorios son la base esencial de la simulación, ya que proporcionan la variabilidad necesaria para modelar situaciones reales y generar resultados diversos.

La calidad de los resultados de una simulación depende en gran medida de la calidad de los números aleatorios generados.

El número no es aleatorio, pero parece serlo, en el sentido en que en una aplicación la relación real entre un número y el siguiente no tiene ningún significado físico.

Los números pseudoaleatorios son generados por medio de una función determinista, no aleatoria y aparentan ser aleatorios. Estos números pseudoaleatorios se generan a partir de un valor inicial aplicando iterativamente la función.

Una sucesión puede considerarse aleatoria si satisface un cierto conjunto de pruebas estadísticas de aleatoriedad.

Se alimenta una función f determinística con una semilla para que produzca un entero x. Este entero se lo vuelve a pasar a f para obtener el siguiente. Si el máximo entero es m, el número aleatorio r es x / m.

Generador Congruencial Lineal

$$x_{n+1} = (ax_n + b) \mod m \qquad 0 \leq x_n < m \qquad \forall n$$

x_0 valor inicial o **semilla** - $x_0 > 0$.

a **elemento multiplicador** - $0 < a < m$.

b **incremento** o **constante aditiva** - $0 <= b < m$.

m **módulo**

El PERÍODO es la subcadena de números xi, dentro de la serie generada, en la que no hay repeticiones de números.

La LONGITUD DEL PERÍODO es el número de elementos distintos en la subcadena.

Cuando la longitud del período coincide con el módulo, se dice que el generador es de CICLO COMPLETO.

GCL Mixto

El incremento o constante aditiva b es diferente de 0.

m debe ser grande (2^32 o 2^64, según la arquitectura de la computadora).

m debe ser potencia p^d , con p siendo la base del sistema (10, 2, 16...) para que el cómputo de m sea eficiente.

Teorema de Hull y Dobell: Periodo completo <=>

- m y b son primos entre sí (no tengan factores comunes excepto el 1)
- q es primo y m % q == 0 => (a - 1) % q == 0
- m % 4 == 0 => (a - 1) % 4 == 0

GCL Multiplicativo

El incremento o constante aditiva b es igual a 0.

Más rápido que GCL Mixto, pero con periodos más cortos.

Los generadores congruenciales multiplicativos NO producen secuencias de CICLO COMPLETO.

Teorema de Hutchinson-Lehmer: Longitud de periodo t = m —-1 con m primo:

- t = m - 1 => m es primo
- m es primo => (m - 1) % t == 0
- (m es primo => t = m - 1) <=> (a <> 0 && ∀ factor p primo de (m - 1): a^((m - 1) / p) % m <> 1)

Generador por Cuadrado Medio

Pasos:

1. Se eleva un valor entero x al cuadrado
2. Se lo completa con ceros para poder determinar una parte central
3. Se toma la parte central
4. El número aleatorio se obtiene como r = x / 10^n, con n siendo el número de dígitos deseados

Algunas desventajas del método: Tiene una fuerte tendencia a degenerar a cero rápidamente. Los números generados pueden repetirse cíclicamente después de una secuencia corta.

Pruebas Estadísticas para validar la aleatoriedad de la sucesión

Las pruebas estadísticas permiten evaluar si los números generados cumplen con las propiedades esperadas de una distribución aleatoria, como uniformidad, independencia y ausencia de sesgos.

Pasar una prueba es una condición necesaria pero no suficiente. Un generador puede pasar una prueba y luego no pasarla si se usa otra semilla u otro segmento del ciclo.

La hipótesis de rechazar o no un generador es el objetivo de estas pruebas estadísticas.

Pruebas de Uniformidad

Evalúan si la distribución de los números generados se ajustan a una distribución uniforme, es decir, si todos los valores tienen la misma probabilidad de aparecer.

Prueba de la Chi Cuadrada

Es una prueba de bondad de ajuste que establece si difiere o no la frecuencia observada de una distribución teórica.

Compara la distribución observada de los números en diferentes intervalos con la distribución esperada bajo la hipótesis de que los números son verdaderamente aleatorios y uniformemente distribuidos.

Hipótesis nula y alternativa:

$$H_0 : FO_i = FE_i$$
$$H_1 : FO_i \neq FE_i$$

H_0 asume que los números generados siguen una distribución uniforme.

Si hay N números aleatorios distribuídos en K intervalos, la FE = N/K y:

$$\chi_0^2 = \sum_{I=1}^{K} \frac{(FO_I - FE_I)2}{FE_I}$$

Se aprueba el test si se satisface que:

$$\chi_0^2 < \chi_{[\alpha;k-1]}^2$$

α: Nivel de significancia: nivel de probabilidad bajo el cual se rechaza la hipótesis nula. Probabilidad de cometer un error al rechazar incorrectamente la hipótesis nula cuando es verdadera. Confianza en los resultados = 1 – α.

k-1: Grados de libertad

La tabla de distribución de Chi Cuadrada proporciona los valores críticos para diferentes niveles de significancia y grados de libertad.

Prueba de Kolmogorov-Smirnov

compara la distribución acumulada empírica con la función de distribución acumulada teórica de una distribución uniforme. Si las diferencias son pequeñas y no significativas, se acepta la hipótesis de que los números aleatorios son uniformes.

FDA (- ∞) = 0
FDA es creciente, de 0 a 1
FDA (∞) = 1

FDA observada: y(i), números aleatorios ordenados de menor a mayor, con i variando de 1 hasta n.

FDA esperada: i/n, siendo n la cantidad de números aleatorios, con i variando de 1 hasta n.

i/n debería ser próximo a y(i). Así, se obtiene la máxima diferencia como MAX |y(i) – i/n|. Este valor se compara con el valor crítico de la tabla de distribución de Kolmogorov-Smirnov D(α, n).

Se aprueba el test si MAX |y(i) – i/n| < D(α, n).

Prueba Serial

Usada para probar uniformidad en dos o más dimensiones.

La prueba SERIAL utiliza el concepto de parejas de números aleatorios que se ubican en k^n celdas para evaluar la uniformidad de la secuencia permitiendo detectar patrones o correlaciones entre subsecuencias adyacentes en un espacio de n dimensiones bajo la organización de números en parejas.

Si se tiene una muestra de tamaño N, se puede construir N/n tuplas de números aleatorios no solapados y contar los puntos que caen en cada celda.

Se esperan (N/n) /k^n tuplas en cada celda.

Se usa la tabla de la Chi Cuadrada. Los grados de libertad son (k^n) – 1.

Se acepta la hipótesis nula si:

$$\chi_0^2 < \chi^2_{[\alpha;k^n-1]}$$

Pruebas de Independencia

Evalúan si cada número generado es completamente independiente de los anteriores.

Para verificar la independencia de los números aleatorios, se usan pruebas estadísticas diseñadas para detectar cualquier correlación o dependencia entre ellos.

Prueba o Test de rachas

Rachas: secuencia de observaciones similares o de eventos de cierto tipo, como de números aleatorios que están todos por encima o por debajo de la media de la serie (0,5 en la uniforme).

Las rachas sobre y bajo la media deberían ocurrir de manera intercalada y de manera impredecible.

Se calcula el número de rachas sobre la media y el número de rachas bajo la media en la secuencia de números aleatorios. Luego, se utiliza una estadística de prueba para evaluar si el número observado de rachas es consistente con lo que se esperaría bajo la hipótesis de aleatoriedad.

n1 y n2: cantidad de números sobre y debajo de la media respectivamente.

N = n1 + n2: longitud total de la secuencia

b: número de rachas

Se calcula:

$$\mu_b = \frac{2n_1n_2}{N} + \frac{1}{2} \qquad \sigma_b^2 = \frac{2n_1n_2(2n_1n_2 - N)}{N^2(N-1)} \qquad Z_0 = \frac{b-\mu_b}{\sigma_b} \sim N(0,1)$$

Si $-Z < Z_0 < Z$, se acepta la hipótesis nula. Z es el valor crítico, sacado de la tabla de la distribución Normal Estándar.

Se puede, además, calcular mediante la tabla: $P(-Z < Z_0 < Z) = P(Z_0 < Z) – P(Z_0 < -Z)$.

Prueba de Poker

Se basa en la frecuencia con que ciertos dígitos se repiten en una serie de números.

Partiendo de la cantidad de dígitos de los números aleatorios a probar, se deben encontrar las posibles clases en que se podrían agrupar según semejanzas y diferencias entre sus dígitos. De cada clase, se debe encontrar su frecuencia relativa esperada (FRE).

Teniendo esto, se puede obtener la frecuencia observada de las clases de una muestra de tamaño N a probar y efectuar una prueba de la Chi Cuadrada entre esta y las FRE * N por cada clase.

Variables Aleatorias

Función matemática que asigna un valor, usualmente numérico, al resultado de un experimento aleatorio

Se obtienen mediante la transformación de distribuciones uniformes básicas en distribuciones más complejas adecuadas para el modelo en cuestión.

Tienen capacidad para:

- Modelar Incertidumbre
- Realizar Análisis de Sensibilidad
- Validar Modelos

Es DISCRETA si los números asignados a los sucesos son puntos aislados. Sus posibles valores constituyen un conjunto finito o infinito numerable.

Es CONTINUA si los valores asignados pueden ser cualesquiera, dentro de ciertos intervalos, es decir, puede tomar cualquier valor de R.

La EXTRACCIÓN de VARIABLES ALEATORIAS implica el uso de algoritmos y técnicas matemáticas para producir números que se comportan de acuerdo a una distribución de probabilidad determinada.

Hay distribuciones teóricas (uniforme, normal, exponencial, Poisson, binomial, geométrica, gamma, beta, lognormal, Weibull, T de Student, chi-cuadrado, F de Fischer-Snedecor) y empíricas.

Una distribución empírica se usa en lugar de una distribución teórica para la extracción de variables aleatorias en simulación cuando se dispone de datos observacionales específicos y se desea que la simulación refleje fielmente las características de esos datos.

Métodos de Extracción de Variables Aleatorias

Procedimiento o algoritmo utilizado para generar valores numéricos que siguen una distribución de probabilidad específica.

- Transformada Inversa: cuando se pueda determinar en forma analítica la FDA
- Aceptación-Rechazo: cuando la cota superior M es pequeña (para minimizar los rechazos)
- Composición: si FDA puede ser expresada como la suma de otras n FDA, y si f puede ser descompuesta como una suma ponderada de otras n densidades
- Convolución: se aplica a dos o más funciones para obtener una nueva

Transformada Inversa

A partir de $f(x)$, se obtiene $u = F(x)$ y se despeja: $x = F^{-1}(u)$. Así, a partir de los números aleatorios u se puede obtener la variable aleatoria x.

Usado para variables continuas y discretas.

Aceptación-Rechazo

Genera variables aleatorias a partir de una distribución deseada utilizando una distribución auxiliar que es más fácil de muestrear.

Para simular valores de f(x) en un intervalo [a, b], se busca una función g(x) más sencilla que envuelva a f(x), y se define M = MAX f(x)/g(x).

La generación de la variable aleatoria se realiza como x* = a + (b-a) * r1, siendo r1 un número aleatorio. Luego, se genera otro número aleatorio r2, y x* se aceptará como variable aleatoria si r2 <= f(x*) / M.

Usado para variables continuas y discretas.

Filas de Espera

Se forma cuando la demanda de un servicio excede la capacidad del sistema para atenderla de inmediato.

La TEORÍA DE COLAS, rama de las matemáticas aplicadas, se dedica a estudiar y analizar los sistemas de espera. A través de modelos matemáticos, se busca entender los patrones de llegada de los clientes, los tiempos de servicio y cómo estos elementos interactúan para formar filas.

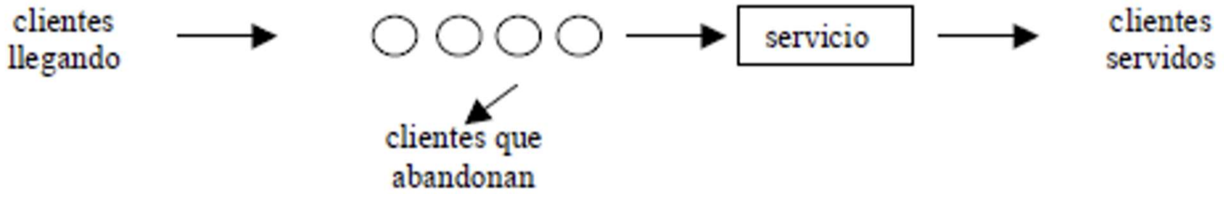
Esta teoría permite predecir y minimizar tiempos de espera, y optimizar el uso de recursos (como tiempos del personal, número de empleados, infraestructura, instalaciones), mejorando así la eficiencia global del sistema.

Razones para el uso de la simulación en la resolución de modelos de fila de espera:

- Complejidad del Sistema
- Flexibilidad y Versatilidad
- Análisis Detallado
- Evaluación de Variabilidad

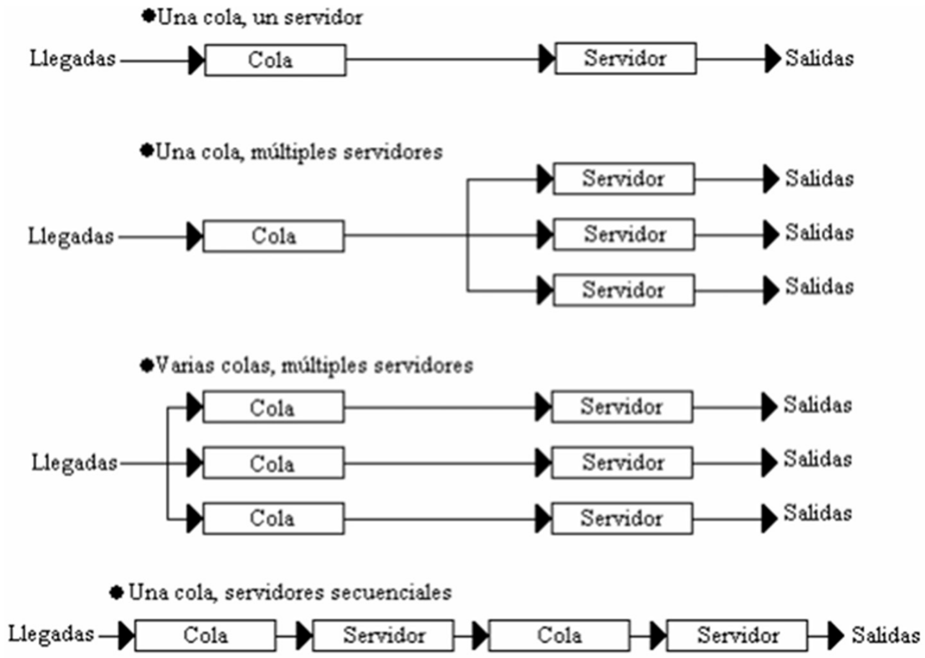
Un SISTEMA DE COLAS se puede describir como: clientes que llegan buscando un servicio, esperan si este no es inmediato, y abandonan el sistema una vez han sido atendidos. En algunos casos se puede admitir que los clientes abandonan el sistema si se cansan de esperar.

Un MODELO DE FILAS DE ESPERA o MODELO DE COLAS en la simulación de sistemas es una representación matemática de un sistema en el que hay entidades (clientes, trabajos, datos, etc.) que llegan para recibir un servicio, pero pueden tener que esperar debido a que los recursos (servidores) están ocupados.



Características

- Patrón de llegada de los clientes. Puede ser determinística o estocástica, por unidad o por lote, estacionario o no (el patrón de llegada varía o no con el tiempo) y los clientes pueden decidir abandonar si pasa mucho tiempo o la cola es muy larga.
- Patrón de servicio de los servidores: describe cómo los clientes son atendidos una vez que ingresan al sistema. Puede ser determinístico o estocástico, por unidad o por lote, estacionario o no, dependiente o no (varía según el número de clientes en la cola).
- Disciplina de la cola: reglas que determinan el orden en que los clientes son atendidos.
    - FIFO
    - LIFO
    - SIRO (Service In Random Order)
    - Prioridades: algunos clientes antes que otros, independientemente de cuándo llegaron.
        - Preemptive: el cliente siendo atendido se va para que se atienda al prioritario
            - El cliente retirado debe volver a empezar
            - El cliente retirado retorna donde se había quedado
        - No preemptive: el prioritario espera a que se termine de atender al cliente actual
- Capacidad del sistema: número máximo de clientes que pueden estar presentes en el sistema simultáneamente. Puede ser limitada o ilimitada.
- Número de canales del servicio: servidores disponibles para atender a los clientes simultáneamente.



- Número de etapas del servicio: número de fases o procesos que un cliente debe atravesar para completar su servicio. Puede ser de una sola o de múltiples etapas (single-stage o multi-stage)

Funcionamiento de una FIFO con un servidor

Por cada cliente, se puede analizar según su tiempo de llegada y los de los anteriores:

- Tiempo de servicio: en que el servidor lo está atendiendo
- Tiempo de ocio del servidor: tiempo desde que se fue el último cliente en la cola antes del actual hasta que este llegó
- Tiempo de espera del cliente: tiempo desde que llegó el cliente hasta que se lo comenzó a atender

Notación de Kendall

A/B/s/K/N/D

- A - Distribución de tiempos entre llegadas
  - M – Poisson (exponencial)
  - D – Determinístico (intervalo constante)
  - G – General (cualquier distribución)
- B - Distribución de tiempos de servicio
  - M – Poisson (exponencial)
  - D – Determinístico (intervalo constante)
  - G – General (cualquier distribución)
- s - Número de servidores
- K - Capacidad del sistema (número o ∞)
- N - Tamaño de la población de clientes (número o ∞)
- D - Disciplina de la cola
  - FIFO
  - LIFO
  - SIRO
  - PR (prioridad)