**Selección de Distribuciones de Probabilidad**

Recolectar Datos

* Simulación Dirigida por los Datos

Se repite lo que pasó en la observación a partir de la cual se obtuvieron los datos. Sirve para validar el modelo.

* Definir una Función Empírica

Se genera una especie de función escalonada. Desventajas: sólo se generan valores entre el mínimo y el máximo que se hayan observado, lo que no significa que no pueda haber valores fuera de ese rango. Se puede generar aleatoriamente, pero tiene limitaciones.

* Determinar una Función/Distribución Teórica a la que se ajustan los Datos

Inferencial.

Parámetros

* De localización: Si cambia, la función se desplaza lateralmente (por el eje x).
* De escala: . Si cambia, se modifica la altura de la función (escala). Se obtiene una función más alargada o más achatada, sin que la localización cambie.
* De forma: . Si cambia, podría ocurrir que una función asimétrica a derecha, pase a ser asimétrica a izquierda, por ejemplo.

Funciones

* Continuas
* Discretas. Solo algunos de los valores del dominio se presentan.

Determinar la Independencia de los Datos

* Gráficos de Dispersión

El área del gráfico es un cuadrado. Se grafican puntos, un valor contra el siguiente: . Hay dependencia cuando los puntos forman una tendencia y no se encuentran dispersos en toda el área.

* Gráficos de Correlación

Puntos de correlación . Ve la relación que existe entre un valor y otro valor que está j lugares separado en la secuencia de datos. Si los datos son independientes, deberían estar cerca de cero. Si se acercan a 1 o -1, hay correlación (positiva o negativa, pero hay); en tal caso, hay que descartar los datos y volver a recolectar.

Los datos deben superar ambas técnicas para poder avanzar.

Pasos para descubrir la función a la que se ajustan los datos

* 1º Paso: Hipotetizar la/s familia/s de Distribución.

1. Estadísticos Descriptivos: moda, media, mediana, varianza, etc.
2. Histogramas y Gráficos de Línea.
3. Resumen de Quantiles y Gráficos de Caja.

* 2º Paso: Estimación de Parámetros (para cada familia candidata)
* 3º Paso: Testear la representatividad de las Distribuciones

1. Procedimientos Heurísticos
   1. Comparación de Frecuencias

Se compara la frecuencia observada y la frecuencia esperada . Lo mismo se puede hacer con

* 1. Gráficos (probabilidad – probabilidad)
  2. Gráficos (quantil – quantil)

1. Procedimientos Estadísticos
   1. Tests.

. IID = Independientes idénticamente distribuidos según una función.

* Test de Chi cuadrado . Trata de armar intervalos equiprobables.
* Test K-D
* Test A-D

**Generador de Números Aleatorios**

* Mecánicos

Preguntar alguien que enuncie un número aleatorio. Depende de la habilidad de ese alguien.

* Tablas
* Métodos Aritméticos

independencia total. secuenciamiento. No son totalmente aleatorios; son pseudoaleatorios.

distribuidos uniformemente entre 0 y 1.

Método de los cuadrados medios

1. Se parte de un número semilla de 4 dígitos.
2. Se lo eleva al cuadrado.
3. Se toman los 4 dígitos del medio.

La bondad del método residía en la semilla. Con una semilla no tan buena, los números se repiten rápidamente.

Características de un buen generador de números aleatorios

* correlación
* Algoritmo rápido y eficiente
* Que se pueda reproducir un stream.
* Streams separados.

Métodos congruenciales lineales

= multiplicador

= incremento

Puede ser que, combinando generadores no tan buenos, se pueda crear un generador bueno.

Testear Generador

* Test Chi-Cuadrado

Se basa en definir un conjunto de intervalos para hallar un valor Chi Cuadrado

N = número de observaciones

K = intervalos

= lo observado, lo que hay en cada intervalo

*Testea tanto uniformidad como independencia.*

* Test Serial

*Solo testea uniformidad.*

Toma los valores generados y arma tuplas de tamaño d (si d = 2, serían pares). Ejemplo:

Siempre en el orden de aparición.

“K” intervalos

Por cada tupla, se fija en qué intervalo se encuentra cada elemento.

*rechazo*

Generación de Variaciones Aleatorias

Generador de Variaciones Aleatorias

* Exacto
* Eficiente
* Robusto (que funcione siempre y con distintos parámetros)
* Complejidad mínima

Métodos:

* Transformada inversa:
  + Parte de la función de distribución.
  + Necesita la inversa de la función De la forma: . .
  + Procedimiento:
    1. Genera
    2. Devolver
  + Ventaja: es rápido.
  + Desventaja: se necesita la inversa de la función (y no todas las funciones tienen).
* Composición:
  + Divide a la función de distribución que no tiene inversa en varias funciones que sí tienen.
  + Procedimiento:
    1. Generar un entero positivo aleatorio . La forma más simple sería generar un número aleatorio, multiplicarlo por la cantidad de funciones, tomar la parte entera y sumarle 1.
    2. Generar
* Convolución
  + de media es una suma de variaciones de funciones exponenciales
  + Procedimiento
    1. Generar
    2. Devolver
* Aceptación y Rechazo
  + Función Mayorizante: . Función tal que
  + Procedimiento
    1. Generar
    2. Generar
    3. Si devolver , sino, volver al paso 1.
  + El secreto está en encontrar una función mayorizante que sea parecida a la que se tiene, para que el método sea más rápido y no deba hacer tantos ciclos antes de cumplir la condición.

Análisis de Datos de Salida

“n” corridas deben ser corridas independientes

Para que sean independientes, deben cumplir 3 condiciones:

* Números aleatorios distintos (mientras no se detenga la simulación, los números aleatorios no se repiten).
* Resetear estadísticos.
* Iguales condiciones iniciales.

Intervalo de Confianza

Obtener mayor precisión

* Error Absoluto: el error se expresa en la misma unidad que el estadístico. Indica un estimativo de la cantidad de corridas que deberían hacerse para obtener el error que se espera.

Puede que indique más corridas de las que en realidad se necesitan, pero asegura que se obtendrá un error que se esté dispuesto a aceptar.

* Error Relativo:

Tipos de Simulación

* Simulaciones Terminales
* Simulaciones No Terminales
  + Parámetros Estacionarios (con transiente inicial, punto l (punto estacionario) y período estacionario)
  + Parámetros Estacionarios cíclicos
  + Otros Parámetros

Para determinar si es un tipo de sistema estacionario:

“n” corridas; “m” datos intermedios

-----------------------------

Promedios móviles: promedios de promedios para suavizar la gráfica y detectar de mejor manera cuándo se estaciona el sistema (punto L).

“w” = ventanas.

Las ventanas van tomando varios “y” y haciendo el promedio. Se grafica y se determina si ya se puede ver con claridad el punto l.

**Comparación de Configuraciones Alternativas**

Siempre se piensa sobre un estadístico que se quiera mejorar.

Estrategias de Comparación

* Todos contra el Estándar
* Todos contra todos

Técnicas de Comparación

* Pares-t
  + Se exige que
  + para
  + Negativo, Negativo: es mejor la opción 1
  + Positivo, Positivo: es mejor la opción 2
  + Negativo, Positivo: no sé cuál es mejor. A veces uno y a veces otro.
* Muestras t de tamaño 2
  + Requiere de independencia entre corridas de los dos sistemas
  + mayor, menor o igual
  + para
  + para

Técnicas de Rankeo y Selección

“K” Sistemas o Configuraciones Alternativas

Tres procedimientos:

* Seleccionar el mejor de los “K” sistemas: mucho esfuerzo de cálculo y procesamiento. Pide bastantes corridas adicionales.
* Seleccionar un conjunto de “m” sistemas que contiene el mejor: menos esfuerzo, pero el conjunto podría contener al peor también. Requiere que luego se comparen todos contra todos para determinar el mejor.
* Seleccionar los “m” mejores sistemas.

Técnicas de Reducción de Varianza

* Variaciones Antitéticas

con

Las corridas antitéticas se hacen trabajando con el complemento de cada número aleatorio usado en la simulación

* + Cov = covariancia (mide la correlación entre dos grupos de números). Es negativa cuando no hay variación.
  + Puede reducir hasta un 30% la variancia original que se hubiera obtenido.
  + Para el intervalo, se utiliza la media de las primeras corridas .
* Variaciones de Control
  + Elegir un estadístico y una variable de control que afecte directamente al mismo (es decir que, si la variable se modifica, el estadístico se ve modificado).
  + Deben guardarse datos de todos los valores de durante la simulación.

**Diseño Experimental y Optimización**

Busca evaluar cómo los cambios en los factores alteran el resultado esperado. Plantear las posibilidades de cambios y obtener la respuesta a los mismos.

Factor:

Ejemplo: cantidad de servidores, política de la cola (FIFO, Prioritaria, etc.).

Pueden ser **controlables** (cantidad de servidores, política de cola) o **no controlables** (cantidad de pedidos).

Pueden ser **cuantitativos** (es un número, contable, medible) o **cualitativos**.

Tienen **niveles**: son los valores que pueden tomar los factores.

Diseño factorial

Por cada factor, hay dos niveles:

* + Estado Original
  + Estado Alternativo, Cambio/Alteración/Modificación (no necesariamente mejorado)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Combinación  Factor | Factor 1 | Factor 2 | Factor 3 | Respuesta |
| 1 | **-** | **-** | **-** |  |
| 2 | **+** | **-** | **-** |  |
| 3 | **-** | **+** | **-** |  |
| 4 | **+** | **+** | **-** |  |
| 5 | **-** | **-** | **+** |  |
| 6 | **+** | **-** | **+** |  |
| 7 | **-** | **+** | **+** |  |
| 8 | **+** | **+** | **+** |  |

Se analizan:

* Efectos Principales (1 por factor):
* Efectos de Interacción

Ejemplo: Sistema de Inventarios

Factores:

* Punto de pedido
  + Nivel Original: 20
  + Nivel Alternativo: 60
* Cantidad de pedidos
  + Nivel Original: 10
  + Nivel Alternativo: 50

R = costo promedio de inventario

|  |  |  |  |
| --- | --- | --- | --- |
| Combinación  Factor | Factor 1 | Factor 2 | Respuesta |
| 1 | **-** | **-** | 144.16 |
| 2 | **+** | **-** | 144.50 |
| 3 | **-** | **+** | 119.99 |
| 4 | **+** | **+** | 147.00 |

* Efectos Principales (1 por factor):

Indica en cuánto se modifica aproximadamente el índice que se está viendo cuando se altera cada factor. En este caso, un valor negativo indica algo bueno, ya que se reduce el costo promedio.

* Efectos de Interacción

**Simulación Continua (Dinámica de Sistemas)**

* También hay cambios y eventos, pero lo que importa es ver el estado del sistema de manera periódica, a **intervalos fijos de tiempo**.
* Se utiliza un mecanismo de avance fijo porque no importa saber cómo cambia el sistema cada vez que ocurre un evento.
* Se identifican **variables**, no eventos (hay eventos, pero no es lo que interesa).
* Las variables llevan asociadas eventos, pero no se identifican de esa forma. Lo que se estudia son las variables.

Aspectos Estructurales

* Límite
  + Variables endógenas (variables que dan información o que tienen un valor y se pueden modificar o se sabe la forma en que se calculan) y exógenas (solo dan valor, pero no se pueden modificar ni se puede cambiar cómo se calculan).
* Elementos: tienen categorías
* Red de Comunicación: conecta elementos del sistema.
  + Línea llena: lo que fluye o pasa, es algo material.
  + Línea de puntos: lo que va por una comunicación así, es información.

Diagrama Causal

Variables:

* Población
* Tasa de natalidad
* Nacimientos
* Tasa de mortalidad
* Muertes
* Tasa de inmigrantes
* Inmigrantes
* Tasa de emigración
* Emigrantes
* Tasa de empleo
* Índice de seguridad

Relaciones: se establecen relaciones (flechas) con + o -, dependiendo de si el cambio de una variable afecta positiva o negativamente en la otra variable.

Luego se identifican **bucles de retroalimentación** (cuando se parte desde una variable y se llega a la misma).

Pueden ser positivos o negativos. Los negativos: sirven para controlar el sistema (si todos fueran positivos, el sistema siempre crecería). Para que haya un bucle negativo, tiene que haber una cantidad impar de relaciones negativas.

Tipos de variables

* Depósitos o Niveles (rectángulos): variable que representa un contenedor. Mantienen el valor cuando la simulación se detiene.
* Variables de Flujo o Válvulas (Forrester: infinito recto con flecha para abajo. Software: círculo con flecha para abajo a la derecha del mismo): regulan o controla las variables de depósito. Las hacen crecer o decrecer.
* Tasas o Parámetros (círculo con una diagonal): ej.: tasa de mortalidad/natalidad, etc.
* Variables Exógenas (círculo con círculo dentro): variables cuyo comportamiento no se puede cambiar.
* Variables Auxiliares (círculo): variables que sirven para hacer cálculos intermedios, no aportan valor en particular.
* Retardos (rectángulo divido en la mitad horizontalmente: arriba, un rectángulo completo; abajo, dividido en 3 cuadrados): tomar información de variables, pero de períodos pasados. Une una variable de flujo con otra (casi siempre).
* Fuentes y Sumideros (nubecita).