Análisis estadístico básico de datos simulados

S. Roncancio 160003741

Facultad de Ciencias Básicas e Ingenierías.
Programa Ingeniería de Sistemas.

Resumen

La realización del análisis estadístico de datos de la simulación de un banco con el proceso de selección aleatoria de tiempo entre llegada de clientes y tiempo promedio de atención al cliente con el uso de un dado de 6 lados y un spinner de 10 segmentos para calcular medidas de desempeño como porcentaje de tiempo de inactividad, el tiempo de espera promedio por cliente, entre otros datos.

El objetivo de la simulación es crear un modelo de un sistema real en nuestro caso del banco para comprender su comportamiento y evaluar estrategias, en el proceso es posible observar que para una mayor exactitud es necesario la mayor cantidad de datos para el análisis.

Palabras clave: Simulación, aleatoriedad, distribución empírica

---

1. Introducción

La motivación del análisis es investigar el comportamiento de la operación de un banco, hasta que se atiende a 20 clientes, y calcular medidas de desempeño como el porcentaje de tiempo de inactividad, el tiempo de espera promedio por cliente, etc. pero para eso es necesario tener algunos conceptos como la definición de simulación por R. E. Shannon es "*La simulación es el proceso de diseñar un modelo de un sistema real y llevar a término experiencias con él, con la finalidad de comprender el comportamiento del sistema o evaluar nuevas estrategias -dentro de los límites impuestos por un cierto criterio o un conjunto de ellos - para el funcionamiento del sistema*" [1]

Aleatoriedad que se asocia a todo proceso cuyo resultado no es previsible más que en razón de la intervención del azar; un ejemplo muy sencillo de un evento aleatorio es el lanzamiento de dados. El resultado de todo suceso aleatorio no puede determinarse en ningún caso antes de que este se produzca.

La función de distribución empírica (FED) o cdf empírica es una función de paso que salta por 1/N a la ocurrencia de cada observación:

$$FDE(x) = \frac{1}{N}\sum_{i=1}^{N} I\{x \leq x_i\}$$

Donde

- $\{A\}$ es el indicador de la función de un evento
- $I\{x \leq x_i\} = \begin{cases} 1 & \text{if } x \leq x_i \\ 0 & \text{if } x > x_i \end{cases}$

El FED estima la verdadera función de densidad acumulativa subyacente de los puntos en la muestra; Se garantiza virtualmente que converge con la distribución verdadera a medida que el tamaño de la muestra se hace lo suficientemente grande.

$$F(x) = \begin{cases} 0 & Si \quad x < 1 \\ \frac{1}{6} & Si \quad 1 \leqslant x < 2 \\ \frac{2}{6} & Si \quad 2 \leqslant x < 3 \\ \frac{3}{6} & Si \quad 3 \leqslant x < 4 \\ \frac{4}{6} & Si \quad 4 \leqslant x < 5 \\ \frac{5}{6} & Si \quad 5 \leqslant x < 6 \\ 1 & Si \quad 6 \geqslant x \end{cases}$$

Función de distribución empírica del lanzamiento del dado de 6 lados teórica

$$F(x) = \begin{cases} 0 & Si \quad x < 1 \\ \frac{1}{10} & Si \quad 1 \leqslant x < 2 \\ \frac{2}{10} & Si \quad 2 \leqslant x < 3 \\ \frac{3}{10} & Si \quad 3 \leqslant x < 4 \\ \frac{4}{10} & Si \quad 4 \leqslant x < 5 \\ \frac{5}{10} & Si \quad 5 \leqslant x < 6 \\ \frac{6}{10} & Si \quad 6 \leqslant x < 7 \\ \frac{7}{10} & Si \quad 7 \leqslant x < 8 \\ \frac{8}{10} & Si \quad 8 \leqslant x < 9 \\ \frac{9}{10} & Si \quad 9 \leqslant x < 10 \\ 1 & Si \quad 10 \geqslant x \end{cases}$$

Función de distribución empírica del spinner de 10 segmentos teórica

Valor medio o también llamada promedio se obtiene a partir de la suma de todos sus valores dividida entre el número de sumandos cumpliendo la siguiente ecuación

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

La desviación estándar es la medida de dispersión más común, que indica qué tan dispersos están los datos con respecto a la media. Mientras mayor sea la desviación estándar, mayor será la dispersión de los datos.

2. Sección experimental

Este laboratorio tiene el fin de observar el comportamiento de la aleatoriedad de números para un análisis básico de datos simulados, para eso se considera el ejercicio planteado "*Example Ad Hoc Simulation*" que consiste en "Considerar la operación de un banco de un solo cajero donde los clientes llegan para el servicio entre 1 y 10 minutos en el tiempo, solo valores enteros, cada valor con la misma probabilidad. Los clientes son atendidos en entre 1 y 6 minutos, también valorados en números enteros e igualmente probables. Restringir los tiempos a valores enteros es una abstracción de la realidad ya que el tiempo es continuo, pero esto ayuda a presentar el ejemplo.

El objetivo es simular la operación del banco, a mano, hasta que se atiende a 20 clientes, y calcular medidas de desempeño como el porcentaje de tiempo de inactividad, el tiempo de espera promedio por cliente, etc. Es cierto que 20 clientes son demasiado pocos para sacar conclusiones sobre el funcionamiento del sistema a largo plazo, pero al seguir este ejemplo, se prepara el escenario para una discusión más detallada en este capítulo y una discusión posterior sobre el uso de la computadora para realizar la simulación. Para simular el proceso, es necesario generar tiempos aleatorios entre llegadas y servicios. Suponga que los tiempos entre llegadas se generan usando una ruleta que tiene posibilidades para los valores del 1 al 10. Suponga además que los tiempos de servicio se generan usando un dado que tiene posibilidades para los valores del 1 al 6."

3. Resultados y análisis

Los resultados que se obtuvieron en la práctica son los siguientes.

$$F(x) = \begin{cases} 0 & Si \quad x < 1 \\ \frac{2}{20} & Si \quad 1 \leqslant x < 2 \\ \frac{6}{20} & Si \quad 2 \leqslant x < 3 \\ \frac{10}{20} & Si \quad 3 \leqslant x < 4 \\ \frac{15}{20} & Si \quad 4 \leqslant x < 5 \\ \frac{18}{20} & Si \quad 5 \leqslant x < 6 \\ 1 & Si \quad 6 \geqslant x \end{cases}$$

Función de distribución empírica del tiempo en servicio (Dado) de 20 clientes

$$F(x) = \begin{cases} 0 & Si \quad x < 1 \\ \frac{3}{19} & Si \quad 1 \leqslant x < 2 \\ \frac{4}{19} & Si \quad 2 \leqslant x < 3 \\ \frac{8}{19} & Si \quad 3 \leqslant x < 4 \\ \frac{10}{19} & Si \quad 4 \leqslant x < 5 \\ \frac{14}{19} & Si \quad 5 \leqslant x < 6 \\ \frac{16}{19} & Si \quad 6 \leqslant x < 7 \\ \frac{18}{19} & Si \quad 7 \leqslant x < 8 \\ \frac{18}{19} & Si \quad 8 \leqslant x < 9 \\ 1 & Si \quad 9 \leqslant x < 10 \\ 1 & Si \quad 10 \geqslant x \end{cases}$$

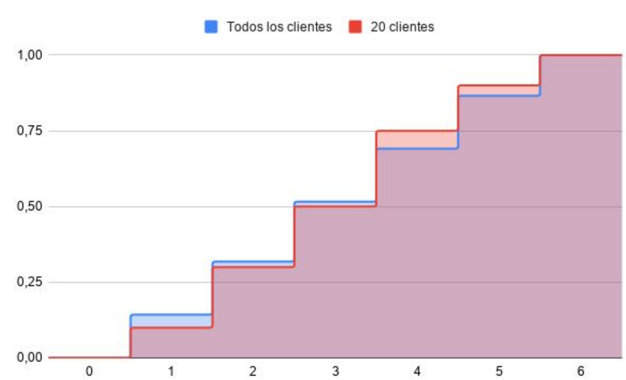
Función de distribución empírica del tiempo entre llegada (spinner) de 20 clientes

$$F(x) = \begin{cases} 0 & Si \quad x < 1 \\ \frac{63}{440} & Si \quad 1 \leqslant x < 2 \\ \frac{140}{440} & Si \quad 2 \leqslant x < 3 \\ \frac{227}{440} & Si \quad 3 \leqslant x < 4 \\ \frac{304}{440} & Si \quad 4 \leqslant x < 5 \\ \frac{381}{440} & Si \quad 5 \leqslant x < 6 \\ 1 & Si \quad 6 \geqslant x \end{cases}$$

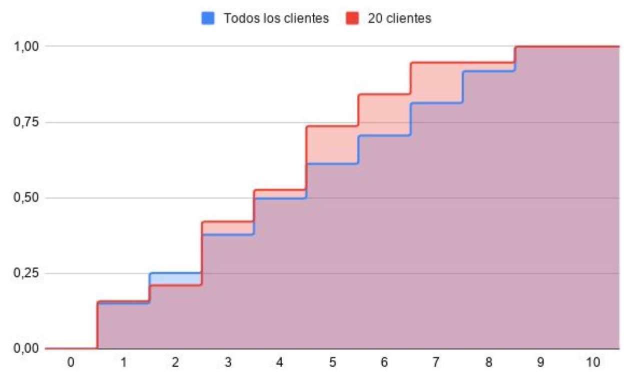
Función de distribución empírica del tiempo en servicio (Dado) de total de clientes

$$F(x) = \begin{cases} 0 & Si \quad x < 1 \\ \frac{63}{418} & Si \quad 1 \leqslant x < 2 \\ \frac{105}{418} & Si \quad 2 \leqslant x < 3 \\ \frac{158}{418} & Si \quad 3 \leqslant x < 4 \\ \frac{208}{418} & Si \quad 4 \leqslant x < 5 \\ \frac{256}{418} & Si \quad 5 \leqslant x < 6 \\ \frac{295}{418} & Si \quad 6 \leqslant x < 7 \\ \frac{340}{418} & Si \quad 7 \leqslant x < 8 \\ \frac{384}{418} & Si \quad 8 \leqslant x < 9 \\ \frac{418}{418} & Si \quad 9 \leqslant x < 10 \\ 1 & Si \quad 1 \geqslant x \end{cases}$$

Función de distribución empírica del tiempo entre llegada (spinner) de total de clientes



Gráfica 1: Gráfica de distribución empírica del dado



Gráfica 2: Gráfica de distribución empírica del spinner

Con el análisis de la gráfica 1 y gráfica 2 se puede observar una gran similitud con los datos en la frecuencia de todos los clientes con la de los 20 clientes y que al mayor volumen de datos la función de distribución experimental tiende a asemejarse a la distribución teórica.

	20 Datos	Todos los datos	Error porcentual
Tiempo medio en el sistema	5.019	4.650	7.4%
Porcentaje de tiempo de inactividad	31.55%	15.85%	49.8%
Tiempo medio de espera por cliente	1.445	1.200	17.0%
Fracción que tiene que esperar	0.538	0.450	16.3%
Tiempo medio de espera de los que se esperaba	3.160	2.660	15.8%

Tabla 1: Diferencias de las medidas de desempeño de la simulación.

Con el análisis de la tabla 1 se puede observar valores altos de diferencia en el error porcentual de los cuales se pueden atribuir al error humano al momento de calcular los valores. Además con los valores calculados es posible afectar al sistema de manera positiva mejorando su eficiencia.

Estudiante	**Average time in system**	**Percent idle time**	**Average waiting time per customer**	**Fraction having to wait**	**Average waiting time of those who waited**
Desviación estándar	1.79	0.13	1.65	0.46	1.62
Valor medio	4.95	0.33	1.39	0.52	3.12

Tabla 2: Medidas de desempeño.

4. Conclusiones

La simulación brinda la oportunidad de probar, diseñar y mejorar procesos y sistemas antes de comprometer el tiempo y los recursos.

Es necesario grandes volúmenes de información en campo para la precisión del modelo.

Referencias.

[1] Shannon, Robert; Johannes, J. D. (n.d.). *Systems simulation: the art and science*. 723–724.