Universidad Nacional Santiago Antúnez de Mayolo
Facultad de Ciencias
Escuela Profesional de Estadística e Informática

COMPUTACIÓN ESTADÍSTICA

M.Sc. Emerson D. Norabuena Figueroa

https://orcid.org/0000-0003-2909-7080

PLANNING

TASK CHECKLIST

- ❑ Algoritmos para generar muestras simuladas
- ❑ Compara métodos de generación

OBJECTIVES

- A

INTRODUCTION

El presente estudio aborda los principales métodos para la generación de variables aleatorias, componente esencial en la simulación estadística, el modelado estocástico y la validación empírica de algoritmos en ciencia de datos, en tanto los procedimientos serán necesarias para la generación de muestras , transformación inversa, método de aceptación y rechazo, y la técnica de composición, destacando sus fundamentos teóricos y criterios de implementación computacional.

El interés de esta radica en contar con mecanismos robustos y reproducibles para simular distribuciones probabilísticas, especialmente en contextos donde la inferencia analítica resulta limitada o inviable. La correcta aplicación de estos métodos permite construir modelos predictivos, evaluar escenarios bajo incertidumbre y optimizar procesos de toma de decisiones basados en datos.[1]

Finalmente, el objetivo proporcionará una guía técnica y didáctica sobre los algoritmos de generación de variables aleatorias, enfatizando su utilidad en entornos de simulación, análisis de Monte Carlo y desarrollo de soluciones automatizadas en plataformas estadísticas como R y Python.

1 Méndez Arias, L. (2022). Métodos de generación de variables aleatorias [Trabajo de fin de grado, Universidad de Zaragoza]. Repositorio Zaguán.

MARCO TEÓRICO

Desde un punto de vista general, los números aleatorios constituyen las entradas de un modelo de simulación y como tales el funcionamiento de la simulación depende de ellos. En tanto, una secuencia de números aleatorios debe tener dos propiedades importantes, uniformidad e independencia.

En consecuencia es necesario precisar que:

- Si (0,1) se divide en n clases o subintervalos de igual longitud, el número esperado de observaciones en cada subintervalo es N/n, donde N es el número total de observaciones.
- La probabilidad de observar un valor en un determinado subintervalo es independiente de las anteriores.

A continuación se presentan características que se debería valorar en cualquier generador de números aleatorios:

Reproductividad: Toda secuencia de números aleatorios debe poderse reproducir con la finalidad de que los experimentos realizados a partir de ella sean contrastables por otro experimentador.
Velocidad: El costo temporal es una variable a considerar, dado que los números aleatorios son tan sólo el inicio del proceso de simulación.
Economía: en cuanto a requerimientos de memoria, por la misma razón que en el caso anterior.
Simplicidad: El generador debe ser suficientemente sencillo de implementar y utilizar.

MARCO TEÓRICO

Algunos métodos de generación:
Existen muchos métodos para la generación de números aleatorios entre 0 y 1, la importancia radica en que los números deben cumplir ciertas características.

Clases de generadores:
- Generadores congruenciales.
- Generadores de registro de desplazamiento.
- Generadores de Fibonacci retardados.
- Generadores no lineales.
- Combinación de generadores.
- Generadores paralelos de números aleatorios.
- Generadores comerciales.

MARCO TEÓRICO

La mayoría de los RNG están basados en congruenciales lineales de la forma:

$$\boldsymbol{x_n = (a_1 x_{n-1} + \cdots + a_k x_{n-k}) \bmod m}$$

donde m se llama *módulo,* $a_1,\ldots, a_k$ son enteros entre −m + 1 y m − 1 llamados multiplicadores con $(a_k \neq 0)$ y k es el orden de la recurrencia. Se puede definir el estado de la recurrencia en el paso n como $S_n = (x_n, \ldots, x_{n+k-1})$. La longitud del periodo máximo es $m^k - 1$.

- **Métodos de congruencias aditivas**

Se precisa de los números $x_1, x_2,\ldots, x_n$. El generador produce una extensión de la secuencia $x_{n+1}, x_{n+2},\ldots$ de la forma siguiente:

$$\boldsymbol{x_i = (x_{i-1} + x_{i-n}) \bmod m}$$

Por definición a = b mod m, si a − b es divisible por m (resto 0). Los número U_i pueden ser generados mediante la ecuación:

$$\boldsymbol{U_i = \frac{x_i}{m}}$$

Por tanto, en el modulo 4, los números 2, 6, 10 y 14 son equivalentes porque (10 − 2), (10 − 6),… son todos divisibles por 4.

Case

Sea una secuencia de enteros dados: x_1, x_2, x_3, x_4, x_5 (58, 35, 90, 95 y 19), generar 3 números pseudoaleatorios entre cero y uno, y m = 100.

La secuencia anterior puede ser ampliada por este método:

$x_6 = (x_5 + x_1) \bmod 100 = 77 \bmod 100 = 77$

$x_7 = (x_6 + x_2) \bmod 100 = 112 \bmod 100 = 12$

$x_8 = (x_7 + x_3) \bmod 100 = 102 \bmod 100 = 2$

Nota técnica:

77 es menor que m y no hay residuo al dividirlo

112 /100 = 1 con residuo 12

Los números aleatorios se obtienen a partir de la relación:

$$\boldsymbol{U_i = \frac{x_i}{m}}$$

para i = n + 1, n + 2, ...

En el caso anterior los números serían:

$$U_1 = \frac{x_6}{100} = 0.77$$

$$U_2 = \frac{x_7}{100} = 0.12$$

$$U_3 = \frac{x_8}{100} = 0.02$$

REFERENCIAS BIBLIOGRÁFICAS

- Méndez Arias, L. (2022). Métodos de generación de variables aleatorias [Trabajo de fin de grado, Universidad de Zaragoza]. Repositorio Zaguán.

OTROS RECURSOS DE INTERÉS

https://www.kdnuggets.com/

http://archive.ics.uci.edu/ml/datasets.php

https://www.cs.ubc.ca/labs/beta/Projects/autoweka/datasets/

https://explodat.cl/Analytics/business-intelligence/la-metodologia-kimball-para-data-warehouses-y-bi-exitosos/

Dive straight in!