

UNIVERSIDADE FEDERAL DE ALAGOAS

INSTITUTO DE COMPUTAÇÃO (IC)

LABORATÓRIO DE COMPUTAÇÃO CIENTÍFICA E ANÁLISE
NUMÉRICA (LACCAN)

Relatório de Pesquisa

Autor:

Marcos G. S. do
Nascimento

Orientador:

Alejandro C. Frery

22/09/2017

1 Levantamento das pesquisas

Para o início da pesquisa referente a minha frente de trabalho que tem por finalidade a implementação de uma biblioteca de funções para simulação e inferência em modelos de dados SAR foram buscadas uma série de boas referências para que fosse construída uma base inicial de conhecimento boa para fornecer suporte às realizações dos objetivos finais do projeto.

Assim, foram acessados por meio da Internet, uma série de artigos de autoria do próprio orientador Alejandro Frery e também de teses de mestrado e doutorado de orientandos deste orientador no tocante à temática da pesquisa.

Como exemplos de artigos e teses acessados e consultados por mim para fins de estudo temos os seguintes artigos:

- The Polarimetric G Distribution for SAR Data Analysis de autoria do professor Alejandro C. Frery, Corina C. Freitas e Antonio H. Correia
- A Model for Extremely Heterogeneous Clutter, Alejandro César Frery et al.
- Um sistema de Análise e Classificação Estatísticas para Imagens SAR de autoria de Pedro Ronalt Vieira, Corina da C. F. Yanasse, Alejandro C. Frery e Sidnei J. S. Sant'Anna

Como exemplos de dissertações de mestrado também vistas, temos:

- Projeto, desenvolvimento e avaliação de classificadores estatísticos pontuais e contextuais para imagens SAR polarimétricas de Antonio H. Correia
- Um novo algoritmo para filtragem de Speckle em imagens SAR de Intensidade baseado em Distâncias Estocásticas de Leonardo M. Torres.

Somado a estes materiais foram buscados também na Internet materiais produzidos por grandes universidades como a PUCRS e a UFPE e, além disso, tutoriais e sites para estudo dos Métodos de geração de números aleatórios e de bibliotecas da plataforma R, entre elas a biblioteca maxLik que se configura como um exemplo interessante do tipo de desenvolvimento que se deseja realizar neste trabalho.

2 A pergunta científica

A pergunta científica de pesquisa foi solicitada ao orientador e este informou que a pergunta que é o norte para o desenvolvimento da minha frente de trabalho é:

“Qual é a melhor forma de gerar observações de variáveis aleatórias GI0?”

Desta pergunta, podemos extrair uma série de indagações do tipo “baseado em qual critério, irei selecionar as possíveis melhores formas de geração de variáveis aleatórias da distribuição G Intensidade Zero?”, entre outras. A expressão “melhor forma” requer definir critérios de qualidade para a posterior classificação das diversas formas de se gerar variáveis aleatórias GI0.

Após receber a pergunta científica, desenvolvi estudos referentes a distribuição GI0 para conhecer o seu comportamento e as suas possíveis aplicações. Além disso, estudei algumas seções referentes à Princípios Gerais de Geração de Variáveis Aleatórias da grande obra de Luc Devroye intitulada “Non-Uniform Random Variate Generation”.

Sabe-se que um dos problemas mais difíceis na geração de variáveis aleatórias é a escolha de um gerador apropriado. Os fatores que desempenham um papel importante nesta escolha incluem velocidade, tempo de configuração inicial (inicialização), gama do conjunto de aplicações, simplicidade e legibilidade, independência de máquina (portabilidade) e tamanho do código compilado. Dessa forma, todos esses critérios devem ser levados em conta para a escolha da forma de geração de variáveis aleatórias da distribuição alvo deste trabalho, a G-Intensidade Zero.

Estudou-se com mais detalhes a distribuição GI0 e foi descoberto que para imagens SAR monoespectrais em intensidade, a distribuição GI0 produz uma boa modelagem para áreas consideradas extremamente heterogêneas. A motivação para a proposta dessa nova classe de distribuições G, em especial a distribuição G0, decorrentes do modelo multiplicativo, é a capacidade que estas distribuições apresentam de modelar áreas extremamente heterogêneas, como áreas urbanas que não podem ser modeladas corretamente com as distribuições K. A vantagem da distribuição G0 se torna evidente através da análise de uma variedade de áreas (floresta urbana, primária e desmatada) de dois sensores.

A partir disso, tendo em mente a motivação da proposta da distribuição que é o foco da pesquisa, foram pesquisadas as características da função de densidade necessárias para uma dada variável aleatória X seguir uma distribuição G-Intensidade Zero (GI0) com parâmetros α, γ e n .

Dessa forma, diz-se que a variável aleatória X possui uma distribuição G-Intensidade Zero (GI0), com os parâmetros definidos acima, denotada por $X \sim GI0(\alpha, \gamma, n)$, supondo-se $\alpha < 0$ e $\gamma > 0$ se a sua densidade, para todo $x \in \mathfrak{R}$, for dada por :

$$f_X(x; \alpha, \gamma, n) = \frac{n^n \Gamma(n - a) x^{n-1}}{\gamma^\alpha \Gamma(n) \Gamma(-a) (\gamma + nx)^{n-\alpha}} \quad (1)$$

Percebe-se com isso a necessidade de estudo também da distribuição Γ visto que a mesma está presente na fórmula da função densidade da distribuição GI0.

Em adição a esses dados pesquisados, estudou-se também o método mais utilizado para geração de números pseudoaleatórios que é o Método Congruente Linear (MCL) também conhecido como método congruente misto. Este método foi primeiramente divulgado em um trabalho desenvolvido pelo Prof. D. H. Lehmer, em 1951, quando dos experimentos executados pelo computador ENIAC no MIT. Em suas pesquisas ele descobriu que restos de sucessivas potências de um número possuem boas características de aleatoriedade. Ele obtinha o n -ésimo número de uma seqüência, tomando o resto da divisão da n -ésima potência de um inteiro a por um outro inteiro m . A partir disso surgiram propostas que são generalizações da proposta de Lehmer e seguem a seguinte fórmula:

$$x_n = ax_{n-1} + b \mod n. \quad (2)$$

Os valores de x_n são inteiros entre 0 e $m - 1$. As constantes a e b são positivas.

A popularidade dos geradores baseados neste método deve-se ao fato de serem facilmente analisados e de algumas garantias de suas propriedades dadas pela teoria das congruências.

Portanto, dentre outros métodos de Geração de Números Pseudoaleatórios, o MCL parece ser o mais adequado para execução da minha frente de trabalho.

3 Evolução da pesquisa

Tendo em vista que a pergunta norte da pesquisa é “Qual é a melhor forma de gerar observações de variáveis aleatórias $GI0$?”, uma grande dificuldade está em como interligar e conectar os conhecimentos pesquisados separadamente a respeito de Métodos de geração de Números Pseudoaleatórios (MCL), os princípios de geração de variáveis aleatórias e as características da distribuição $GI0$ (função densidade, parâmetros, etc).

Como visto que as informações em separado já foram buscadas e descritas no presente relatório, o maior desafio nessa etapa seguinte consiste em como interligar e combinar as informações encontradas para se chegar no objetivo final da pergunta científica.

Em outras palavras, como mencionado anteriormente, cada um desses tópicos foi pesquisado e absorveu-se um conhecimento sobre cada uma dessas linhas de pesquisa, o desafio, agora, se mostra em como agregar isso de forma lógica e eficiente para se atingir o objetivo final da pergunta científica descrita acima e assim trilhar no caminho certo rumo ao objetivo final da minha pesquisa.