Generación de funciones de densidad de probabilidad ajustadas a mediciones de variables eléctricas:

caso de Rayleigh y Gamma

Rodrigo López Soto

Escuela de Ingeniería Eléctrica
Universidad de Costa Rica

Proyecto Eléctrico, Agosto 2021

Tabla de Contenidos

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Tabla de Contenidos

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Justificación

El estudio del modelo matemático realizado permite responder la pregunta: ¿Cuál es un modelo adecuado para describir el comportamiento de un conjunto finito de mediciones eléctricas mediante mezclas de funciones de densidad de Rayleigh y Gamma? Para ello, se requiere:

- Utilizar modelos estadísticos como herramienta de estudio de fenómenos eléctricos
- Realizar la limpieza de los datos obtenidos para seleccionar las mediciones de interés.
- Verificar que el modelo propuesto se ajusta al comportamiento real del conjunto de mediciones.

Objetivo general

Diseñar un algoritmo para la generación de funciones de densidad de probabilidad no gaussianas ajustadas a mediciones de variables eléctricas, con base en las funciones de densidad Rayleigh y Gamma.

Objetivos específicos

1. Determinar las características principales en las mediciones de las variables eléctricas para las cuales se implementará el algoritmo.
2. Obtener el algoritmo utilizando la combinación de funciones de densidad de probabilidad no gaussianas, específicamente las funciones de densidad Rayleigh y Gamma.
3. Validar el algoritmo mediante pruebas basadas en los datos reales obtenidos de la Unidad de Verificación de la Calidad del Suministro Eléctrico de la Universidad de Costa Rica y comparar los resultados obtenidos de la validación del algoritmo con los resultados de otras familias de funciones de densidad probabilística utilizadas en estudios previos.

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Alcances

Se pretende

- Generar funciones de mejor ajuste en los casos de Rayleigh y Gamma
- Evaluar el comportamiento del modelo con mediciones reales

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Metodología

1. Revisión bibliográfica sobre el algoritmo de esperanza-maximización.
2. Estudio del entorno de Matlab que se usará para la implementación del algoritmo.
3. Diseño del algoritmo para utilizar las funciones de densidad de probabilidad no gaussianas.
4. Implementación del algoritmo diseñado utilizando Matlab.
5. Validación del algoritmo diseñado con datos reales obtenidos de la UVECASE.
6. Comparación de los resultados obtenidos entre las funciones estudiadas y otra familia de funciones presentadas en trabajos previos.

EIE
Escuela de Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Tabla de Contenidos

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Verosimilitud

Definición:

Sea $x = (x_1, x_2, ..., x_n)^T$ un conjunto de observaciones independientes, y f una función de densidad probabilística con parámetros dados por Θ, se define verosimilitud como:

$$L(\Theta|x) = f(x|\Theta) = \prod_{i=1}^{n} f(x_i|\Theta) \tag{1}$$

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Log-verosimilitud

En ocasiones por facilidad en los cálculos se trabaja en escala logarítmica.

Definición:

Sea $x = (x_1, x_2, ..., x_n)^T$ un conjunto de observaciones independientes, y f una función de densidad probabilística con parámetros dados por Θ, se define log-verosimilitud como:

$$\log(L(\Theta|x)) = \log(f(x|\Theta)) = \sum_{i=1}^{n} \log f(x_i|\Theta) \tag{2}$$

EIE
Escuela de Ingeniería Eléctrica

Universidad de Costa Rica

Permite maximizar log-verosimilitud de una muestra de datos incompleta.

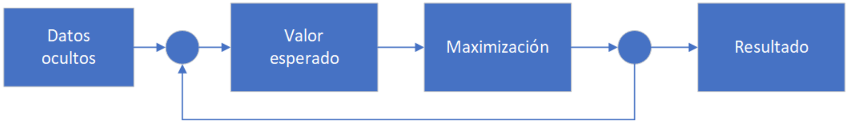

Figura 1: Algoritmo esperanza-maximización. Autoría propia.

Algoritmo Esperanza-Maximización (EM)

- Permite la obtención de la esperanza y los valores que la maximizan
- La maximización en la esperanza implica maximización en log-verosimilitud.
- El algoritmo ubica un máximo local, pero no garantiza que sea el máximo global.
- La maximización se realizó utilizando multiplicadores de Lagrange

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Distribución de Rayleigh

La distribución Rayleigh es encontrada en diversa cantidad de fenómenos. Es una distribución cuya PDF normalizada viene dada por la ecuación 3.

$$f_x(x, \sigma) = \frac{x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}}; x > 0 \tag{3}$$

Donde σ es el parámetro a ajustar.

EIE
Escuela de Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

La PDF de la distribución Gamma viene dada por la ecuación 4.

$$f_x(x, k, \beta) = \frac{\beta^k}{\Gamma(k)} x^{k-1} e^{-\beta x}; x > 0 \tag{4}$$

Donde k y β son parámetros que que se ajustan por medio del algoritmo.

Tabla de Contenidos

EIE Escuela de Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Datos ocultos

- Se supondrá que cada observación x_i es generada por un único componente de f, f_j.
- Los datos ocultos se tratarán como una familia de indicadores y_{ij} que es 1 si el componente i fue el que generó la observación x_j, y 0 de lo contrario (como una delta de Kronecker).
- Estos datos serían generados por una variable aleatoria Y con una función de distribución de probabilidad $g(y)$.

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Esperanza

Los datos ocultos permiten obtener la probabilidad de escoger la componente f_n:

$$E_y\left[y_{nm}|x,\Theta^{(t)}\right] = \frac{\alpha_n^{(t)} f_n(x_m|\theta_n^{(t)})}{\sum_{k=1}^{N}\alpha_k^{(t)} f_k(x_i|\theta_k^{(t)})} \tag{5}$$

Con esto, la esperanza se calcula de la siguiente manera:

$$Q(\Theta|\Theta^t) = \sum_{m=1}^{M}\sum_{n=1}^{N} \frac{\alpha_n^{(t)} f_n(x_m|\theta_n^{(t)})}{\sum_{k=1}^{N}\alpha_k^{(t)} f_n(x_i|\theta_k^{(t)})} \log \alpha_n f_n(x_m|\theta_n) \tag{6}$$

EIE
Escuela de Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Maximización de pesos

Para el caso de los pesos se resuelve mediante multiplicadores de Lagrange el problema:

$$\begin{cases} Q(\Theta, \Theta^t), & \text{Maximización} \\ \sum_{n=1}^{N} \alpha_n - 1 = 0, & \text{Restricción} \end{cases} \tag{7}$$

Se obtiene:

$$\alpha_i = \frac{1}{M} \sum_{m=1}^{M} \left(\frac{\alpha_i^{(t)} f_i(x_m|\theta_i^{(t)})}{\sum_{k=1}^{N} \alpha_k^{(t)} f_n(x_i|\theta_k^{(t)})} \right) = \frac{1}{M} \sum_{m=1}^{M} E_y \left[y_{im}|x, \Theta^{(t)} \right] \tag{8}$$

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Maximización de parámetros de distribuciones individuales

El problema de optimización se analizará para cada función de distribución de probabilidad, de la siguiente manera:

$$
\begin{aligned}
&\frac{\partial}{\partial \theta_i} Q(\Theta, \Theta^t) = 0 \\
\iff &\frac{\partial}{\partial \theta_i} \sum_{m=1}^{M} \sum_{n=1}^{N} \frac{\alpha_n^{(t)} f_n(x_m|\theta_n^{(t)})}{\sum_{k=1}^{N} \alpha_k^{(t)} f_n(x_m|\theta_k^{(t)})} \log \alpha_n f_n(x_m|\theta_n) = 0 \\
\iff &\sum_{m=1}^{M} \frac{\alpha_i^{(t)} f_i(x_m|\theta_i^{(t)})}{\sum_{k=1}^{N} \alpha_k^{(t)} f_n(x_m|\theta_k^{(t)})} \frac{\partial}{\partial \theta_i} (\log \alpha_i f_i(x_m|\theta_i)) = 0
\end{aligned}
\tag{9}
$$

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE
COSTA RICA

Maximización Rayleigh

Se tiene un solo parámetro, que se optimiza de la siguiente manera:

$$\sigma_i^2 = \frac{\sum_{m=1}^{M} E_y \left[y_{im} | x, \Theta^{(t)} \right] x_m^2}{2 \sum_{m=1}^{M} E_y \left[y_{im} | x, \Theta^{(t)} \right]} \tag{10}$$

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Maximización Gamma

Se utiliza la aproximación [2]:

$$\log(\Gamma(k_i)) \approx (k_i - \frac{1}{2}) \log(k_i) - k_i + \frac{1}{2} \log(2\pi) \tag{11}$$

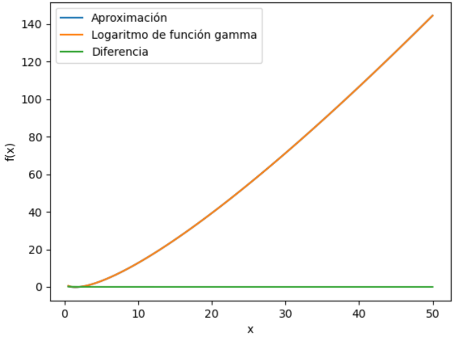

Figura 2: Aproximación utilizada.

Maximización Gamma

Con esto, se obtiene:

$$\frac{k_i}{\beta_i} = \frac{\sum_{m=1}^{M} E_y\left[y_{im}|x, \Theta^{(t)}\right] x_m}{\sum_{m=1}^{M} E_y\left[y_{im}|x, \Theta^{(t)}\right]} \tag{12}$$

$$k_i = \frac{\sum_{m=1}^{M} E_y\left[y_{im}|x, \Theta^{(t)}\right]}{\sum_{m=1}^{M} E_y\left[y_{im}|x, \Theta^{(t)}\right]\left(\log\left(\frac{k_i}{\beta_i}\right) - \log(x_m)\right)} \tag{13}$$

$$\beta_i = \frac{k_i}{\frac{\sum_{m=1}^{M} E_y\left[y_{im}|x,\Theta^{(t)}\right] x_m}{\sum_{m=1}^{M} E_y\left[y_{im}|x,\Theta^{(t)}\right]}} \tag{14}$$

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Parámetros iniciales Rayleigh

$$\alpha_1^{(0)} = \ldots = \alpha_1^{(N)} = \frac{1}{N} \tag{15}$$

$$\sigma = \frac{\mu}{\sqrt{\frac{\pi}{2}}} \tag{16}$$

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Parámetros iniciales Gamma

$$\alpha_1^{(0)} = \ldots = \alpha_1^{(N)} = \frac{1}{N} \tag{17}$$

$$\beta = \frac{\mu}{\sigma^2} \tag{18}$$

$$k = \mu\beta = \frac{\mu^2}{\sigma^2} \tag{19}$$

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Resumen de los algoritmos diseñados

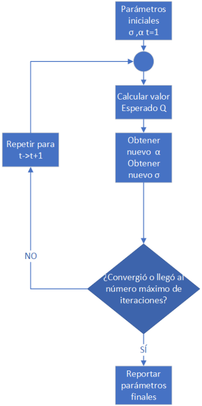

Figura 3: Algoritmo Rayleigh.

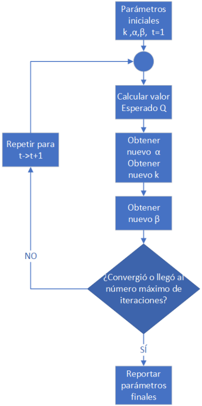

Figura 4: Algoritmo Gamma.

EIE Escuela de Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Tabla de Contenidos

EIE
Escuela de Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Cuadro 1: Promedios obtenidos para el conjunto de datos

Fase	Promedio impares/(%)	Promedio pares/(%)
Conjunto completo	0.2496	0.0275
C-A	0.2595	0.0270
B-C	0.2443	0.0290
A-B	0.2450	0.0265

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Ejemplos de histogramas para fase BC

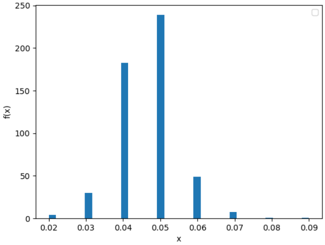

Figura 5: Histograma de cuarta armónica de fase BC.

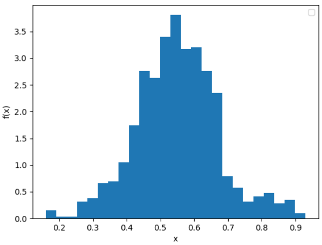

Figura 6: Histograma de séptima armónica de fase BC.

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Resultados: caso quinceava armónica de fase BC

PDF	Log-verosimilitud
Rayleigh	1692.12
Gamma	1741.61
Log-normal	1748.60 [1]
Normal	1735.96

Cuadro 2: Resultados para quinceava armónica de la fase BC

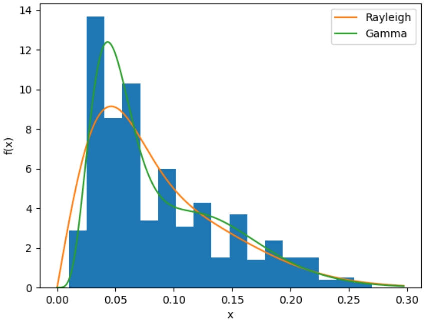

Figura 7: Resultados para quinceava armónica de fase BC.

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Resultados: caso quinta armónica de fase CA

PDF	Log-verosimilitud
Rayleigh	-84.09
Gamma	579.96
Log-normal	583.38 [1]
Normal	585.89

Cuadro 3: Resultados para quinta armónica de la fase CA

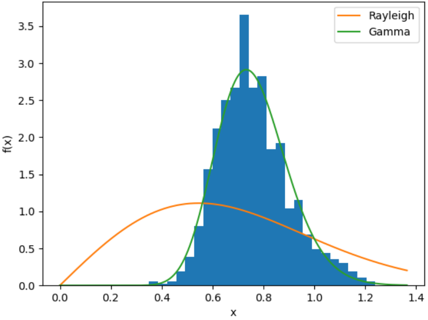

Figura 8: Resultados para quinta armónica de fase CA.

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Resultados: caso tercera armónica de fase BC

PDF	Log-verosimilitud
Rayleigh	66.25
Gamma	496.36
Log-normal	497.37 [1]
Normal	491.43

Cuadro 4: Resultados para tercera armónica de la fase BC

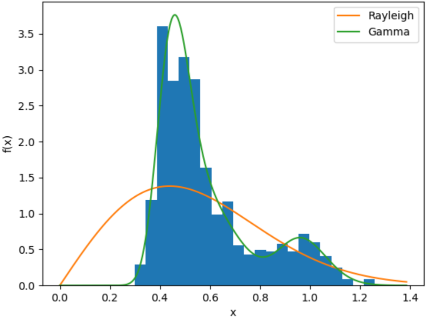

Figura 9: Resultados para tercera armónica de fase BC.

EIE
Escuela de Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Tabla de Contenidos

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Conclusiones

- Se logró diseñar un algoritmo para la generación de funciones de densidad de probabilidad no gaussianas ajustadas a mediciones de tensión RMS, con base en las funciones de densidad Rayleigh y Gamma.
- Se determinó las principales características de las mediciones tensión RMS correspondientes a los armónicos del 2 al 25 para una muestra de datos de la última semana de marzo de 2017, obtenidos de la Unidad de Verificación de la Calidad del Suministro Eléctrico (UVECASE) de la Universidad de Costa Rica.

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Conclusiones

- Se obtuvo un algoritmo utilizando una combinación de funciones de distribución de probabilidad no gaussianas, específicamente las de Rayleigh y Gamma
- Se validó el algoritmo diseñado mediante pruebas en datos reales obtenidos de la Unidad de Verificación de la Calidad del Suministro Eléctrico (UVECASE) de la Universidad de Costa Rica.

EIE
Escuela de Ingeniería Eléctrica

Conclusiones

- Se determinó que en 67 de las 72 variables analizadas el algoritmo utilizando distribución Gamma convergió, y dio mejores resultados que utilizando distribuciones de Rayleigh en estos casos.
- Se observó que todos los datos utilizados fueron positivos, sin embargo, en algunas de las variables las mediciones no inician exactamente en cero, de manera que para el caso de la distribución de Rayleigh los resultados obtenidos no fueron los más adecuados.

Recomendaciones

- Problemas con Rayleigh (sustituir por Weibull).
- Se recomienda comparar los resultados obtenidos con otros algoritmos que se han implementado con los obtenidos utilizando la clase *GaussianMixtures* de la librería *SkLearn* de *Python*.
- Se sugiere seleccionar el número de componentes utilizando algún criterio científico como el criterio de información de Akaike (AIC).

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

¡Muchas Gracias!

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA

Referencias

Oscar Arguedas.
Generación de funciones de densidad de probabilidad ajustadas a mediciones de diversas variables físicas.
Proyecto Eléctrico, 1, 2018.

Edmund Taylor Whittaker and George Neville Watson.
A course of modern analysis.
Courier Dover Publications, 2020.

EIE
Escuela de
Ingeniería Eléctrica

UNIVERSIDAD DE COSTA RICA