

1 Un proceso de fábrica de software tiene una especificación de 30-38 bugs por cada 1,000 líneas de código. Durante un muestreo aleatorio Ud. obtuvo los siguientes valores:

32,33,35,30,30,34,32,31,33,35,32,31,33,32,27,35,33,24,36,35,32,33,29,36,32,33,32,34,33,37,35,34,33,30,34,31,34,33,32,35,32,36,34,30,33,31,36,34,31,31,34,36,35.

Comente acerca de la capacidad del proceso de cumplir la especificación.

2 Visite el sitio "*Chronology of Data Breaches Security Breaches 2005 – Present*" (<http://www.privacyrights.org/data-breach>). Este archivo documenta diferentes problemas de seguridad en el mundo desde el 2005. Descargue el archivo CSV y procéselo con Excel®.

Teniendo en cuenta sólo las fechas construya una "Moving Range Control Chart – XmR" con Count y MR Values. Grafique también los UCL y LCL.

1- ¿Qué conclusiones puede obtener de las gráficas? ¿Es este proceso en "steady" state?

2- ¿Hay variación en sus resultados si "Michael Caine, Morgan Freeman, and Liam Neeson walk into a bar..." y Usted decide usar mejor la mediana?

3- ¿Si Ud. Decidiera agrupar en 'cajas' con la regla de Sturges ¿cambiarían sus resultados?

4- ¿Se trata más bien de un proceso caótico?

3 Uno de los primeros estudios acerca de procesos caóticos encontró una serie de ecuaciones de apariencia anodina que sin embargo describen bastante bien el ascenso y ocaso de poblaciones completas en nichos ecológicos.

La ecuación se conoce como "*logistic map*" <http://mathworld.wolfram.com/LogisticMap.html> y de forma muy elegante se expresa como la ecuación en diferencias:

$$x_{n+1} = \lambda x_n(1 - x_n)$$

Es decir, la población en un momento dado (tiempo $n+1$) es una función de la población en el período de tiempo inmediatamente anterior (n) con una constante "Malthusiana" de reproducción (λ) que indica que tan prolífica es una especie.

El siguiente código sagemath presenta esta función de una manera sencilla de operar.

```
@interact

def_( tasa=slider(0, 4.2, step_size=0.01, default=2.7, label='Lambda'), x0 = slider(0.0, 1.0, step_size=0.01,
default=0.02, label='Xo'), ancho=slider(0, 500, step_size=1, default=30, label='Width') ):
    x = x0
    L = line( [(0,0),(ancho,0)], color="black", thickness=4 )
    for j in range(ancho):
        xNew = tasa*x*(1-x)
        L = L + line( [(j, x), (j+1, xNew)], color="red", thickness=2 )
        x = xNew

    show (L)
```

Nótese el comportamiento de la gráfica dependiendo de la constante Malthusiana, especialmente en los rangos [0.0-0.99], [1.0-2.99], [3.0-3.44], [3.45-3.54], [3.55-3.56], [3.57-3.82], [3.83-3.99] y [4.00-..] ¿qué observa?

Juegue también con diferentes valores iniciales (x_0)... Nótese la extrema sensibilidad del resultado final de este valor inicial. La verdad no debería sorprenderlo, la expresión para x_3 tiene la apariencia:

$$x_3 = \lambda^3(1 - x_0)x_0(1 - \lambda x_0 + \lambda x_0^2)(1 - \lambda^2 x_0 + \lambda^2 x_0^2 + \lambda^3 x_0^2 - 2\lambda^3 x_0^3 + \lambda^3 x_0^4)$$

Por lo que un pequeño cambio en el parámetro x_0 (“*It has been said that something as small as the flutter of a butterfly's wing....*”) causa un cambio grande difícilmente predecible en x_3 (“*... can ultimately cause a typhoon halfway around the world*”).

¿hay forma de ajustar los parámetros λ y x_0 para que refleje el comportamiento de la “*Chronology of Data Breaches Security Breaches 2005 – Present*”.

4 Este ejercicio propone un trabajo en el lenguaje de programación científica GNU Octave (<https://www.gnu.org/software/octave/>) de un ejercicio de “*Anomaly Detection*” en *Machine Learning*. Como requisito inicial Usted debe descargar la aplicación GNU Octave y familiarizarse con su funcionamiento. Un buen tutorial, - quizás demasiado extensivo- lo podrá encontrar por ejemplo en <https://www.gnu.org/software/octave/doc/interpreter/>.

Adjunto al taller encontrará el archivo ‘xfile.mat’, un archivo con 307 parejas de registro de valores tomadas de ODDS: “*A large collection of publicly available outlier detection datasets with ground truth in different domains*”

(<http://odds.cs.stonybrook.edu>) que corresponden a registros de uso de procesador y consumo de ancho de banda de un servidor Internet.

En Octave el comando

```
load('Xfile.mat');
```

carga en la variable X los contenidos de este archivo, de manera que después el archivo X almacenará una matriz de 307x2 elementos.

A Para la primera parte del taller Usted supondrá que las variables uso de procesador y consumo de ancho de banda son independientes y basado en la información elaborará un modelo de “*unsupervised learning*”.

- ¿Cuál es el valor de $\mu_1, \sigma_1, \mu_2, \sigma_2$?
- Si $\epsilon=8.990853 \times 10^{-5}$, calcule la probabilidad para la muestra (19,10) ¿será normal o anómala?
- Calcule ahora la probabilidad para la muestra (17,12) con el mismo valor de ϵ , ¿será normal o anómala?

B El lector cuidadoso habrá notado que las variables uso de procesador y consumo de ancho de banda en la vida real son un tanto dependientes para una aplicación web típica. Construya ahora un modelo de “*unsupervised learning*” en el cual quepa la posibilidad que las variables tengan algún grado de correlación.

- ¿Cuál es el valor de las matrices μ de promedios y Σ de covarianza? ¿coinciden estos valores con lo esperado?
- Si $\epsilon=9.06577 \times 10^{-5}$, calcule la probabilidad para la muestra (19,10) ¿será la muestra normal o anómala?
- Calcule ahora la probabilidad para la muestra (17,12) con el mismo valor de ϵ , ¿será normal o anómala?