

Benemérita Universidad Autónoma de Puebla
Minería de Datos
Moisés Machorro Portilla
202232944

Manual de Usuario:

Simulador de Distribuciones Estadísticas



Introducción

Bienvenido al Simulador de Distribuciones Estadísticas, esta herramienta tiene el objetivo de proporcionar una plataforma visual e interactiva para comprender, simular y analizar diversas distribuciones de probabilidad, así como algoritmos fundamentales de la minería de datos como el de Maximización de la Esperanza (EM). El programa permite generar muestras aleatorias, visualizar su distribución a través de histogramas y comparar los datos simulados con las funciones teóricas.

Obtención del código

Para utilizar el simulador, primero necesitas descargar el código fuente desde el repositorio oficial en GitHub.

Repositorio: https://github.com/ELMOOY/function_simulator

Existen dos métodos principales para obtener el código:

Método 1: Descargar el archivo ZIP (Recomendado)

Esta es la forma más sencilla si no tienes Git instalado.

1. Abre tu navegador web y ve a la dirección del repositorio.
2. Haz clic en el botón verde que dice "< > Code".
3. En el menú desplegable, selecciona la opción "Download ZIP".
4. Guarda el archivo function_simulator-main.zip en tu computadora y descomprímelo en una carpeta de tu elección.

Método 2: Clonar con Git

Si tienes Git instalado en tu sistema, puedes clonar el repositorio, abre una terminal y ejecuta el siguiente comando:

```
git clone [https://github.com/ELMOOY/function_simulator.git]  
(https://github.com/ELMOOY/function_simulator.git)
```

Esto creará una carpeta llamada **function_simulator** con todo el código del proyecto.

Requisitos de instalación

Para ejecutar el simulador, es necesario tener Python instalado en tu sistema y las siguientes librerías:

- NumPy: Para cálculos numéricos.
- Pandas: Para la manipulación y carga de datos (incluyendo archivos .csv y .xlsx).
- Matplotlib: Para la generación de todos los gráficos 2D y 3D.
- SciPy: Utilizado para funciones estadísticas (ej. stats.norm).
- Pillow (PIL): Para cargar y mostrar las imágenes guardadas in la galería del comparador.
- Openpyxl: Requerido por Pandas para poder leer archivos de Excel (.xlsx).

Puedes instalarlas todas juntas ejecutando el siguiente comando en tu terminal:

```
pip install numpy pandas matplotlib scipy pillow openpyxl
```

Inicio del programa

Para iniciar la aplicación, abre una terminal, navega hasta el directorio donde descargaste y descomprimiste el proyecto y ejecuta el siguiente comando:

```
python simuladorFunciones.py
```

la aplicación se abrirá en modo de pantalla completa, mostrando el Menú Principal.



Menú principal

Al iniciar la aplicación, se te presentará un menú principal a pantalla completa, este menú es el punto de partida para todas las funciones del programa.

Las opciones disponibles son:

- Simuladores (Columna Izquierda y Derecha):
 - Bernoulli
 - Binomial
 - Exponencial
 - Normal Univariada
 - Normal Bivariada (Gibbs)
 - Función Particular
- Módulo de Análisis (Columna Derecha):
 - Algoritmo EM (Cáncer)
 - Comparador de Gráficas EM
- Salir:
 - Cierra la aplicación.

Simuladores de Distribuciones

Estándar

Esta sección cubre todas las opciones de simulación excepto el "Algoritmo EM".

¿Cómo usar un simulador?

1. Selecciona una distribución: Haz clic en el botón de la distribución que deseas simular (ej. "Normal Univariada").
2. Ingresa los Parámetros: Serás llevado a la ventana de simulación, en el panel de control de la izquierda, encontrarás:
 - a. Descripción: Una breve explicación de la distribución y sus parámetros.
 - b. Campos de Parámetros: Ingresa los valores requeridos (ej. "Media (μ)" y "Desv. Estándar (σ)" para la Normal).
 - c. Tamaño muestra: Define cuántos puntos de datos deseas generar (ej. 1000).

Simuladores de Distribuciones Estándar

¿Cómo usar un simulador?

3. Ejecuta la Simulación:

- Botón "Simular": Genera los datos y muestra un histograma (para 1D) o un gráfico de dispersión (para 2D) en el panel de visualización de la derecha.
- Botón "Superponer Función": Dibuja la función teórica (PDF, PMF o contornos) sobre el gráfico generado, esto te permite comparar tus datos simulados con la distribución real.
- Botón "Visualizar en 3D" (Solo para Normal Bivariada y Función Particular): Abre una nueva ventana que muestra la superficie de densidad teórica junto con la muestra de datos generada.

4. Paneles de la Ventana de Simulación

- Panel de Control (Izquierda):
 - Contiene todos los campos de entrada, descripciones y botones de acción.
 - Resultados: Para distribuciones 1D, muestra un resumen de los datos generados. Para la normal bivariada, activa un botón para ver la muestra completa en una ventana nueva.
 - Botón "Limpiar": Borra todos los campos de entrada y el gráfico.
 - Botón "Regresar al Menú": Cierra el simulador y vuelve al menú principal.
- Panel de visualización (Derecha):
 - Muestra el gráfico (histograma o gráfico de dispersión) generado por matplotlib.

Módulo: Algoritmo EM (Cáncer)

Está diseñado para cargar un conjunto de datos (como el de cáncer) y aplicar el algoritmo EM para encontrar los parámetros de una mezcla de dos distribuciones normales que mejor se ajusten a los datos.

Este módulo te permite ver visualmente cómo el algoritmo "aprende" los parámetros correctos iteración por iteración.

Pasos de uso detallados

Paso 1: Cargar los Datos

1. Haz clic en el botón "1. Cargar Archivo" en el panel izquierdo.
2. Se abrirá una ventana de explorador de archivos.
3. Navega hasta la ubicación de tu archivo de datos y selecciónalo, el programa soporta archivos de Excel (.xlsx) y CSV (.csv).

Paso 2: Seleccionar la Variable

1. Una vez cargado el archivo, el menú desplegable "2. Variable" se activará (dejará de estar gris).
2. Haz clic en él. Verás una lista de todas las columnas numéricas que se encontraron en tu archivo.
3. Selecciona la variable que deseas analizar (por ejemplo, 'radius (nucA)', 'PC1', 'texture (nucA)', etc.).
4. Inmediatamente, verás el histograma de esa variable en el panel de visualización de la derecha. Esto te da una vista previa de la distribución de tus datos.

Paso 3: Configurar y Ejecutar el Algoritmo

1. Localiza el campo de texto "Número de Iteraciones".
2. Escribe cuántos pasos de "aprendizaje" deseas que realice el algoritmo, Un número bajo (ej. 2 o 3) te permitirá ver los cambios iniciales, pero un número más alto (ej. 10 o 20) mostrará el resultado más ajustado.
3. Haz clic en el botón "3. Ejecutar EM".
4. El programa realizará los cálculos. En segundo plano, está estimando las medias (μ), desviaciones estándar (σ) y pesos (π) de dos distribuciones normales que, combinadas, se ajustan a tu histograma.

Módulo: Algoritmo EM (Cáncer)

Análisis de Resultados (Panel Izquierdo)

Al terminar la ejecución, el panel izquierdo se llenará de nuevas opciones de análisis, aquí es donde ocurre todo:

Análisis Iteración por Iteración (Checklist)

Aparecerá una nueva sección con una lista de casillas de verificación para cada iteración que ejecutaste (ej. "Iter 1", "Iter 2", ...), para cada iteración, puedes activar:

1. N1: Muestra la curva de la primera componente normal (generalmente de color azul/cyan).
2. N2: Muestra la curva de la segunda componente normal (generalmente de color rojo/magenta).
3. Mezcla: Muestra la curva combinada (la suma ponderada de N1 y N2) para esa iteración.
4. Val: Muestra los valores de los parámetros (μ , σ , π) de esa iteración directamente sobre la gráfica.

Consejo: Prueba activando "N1", "N2" y "Val" solo para "Iter 1", Luego desactívalos y activa los mismos para "Iter 5", verás cómo las curvas y los valores han cambiado (evolucionado) para ajustarse mejor a los datos.

Módulo: Algoritmo EM (Cáncer)

Análisis de Resultados (Panel Izquierdo)

Controles Principales (Botones)

Debajo de los botones de ejecución, tendrás más herramientas:

- Botón "Superponer Mezcla Final":
 - Este es un interruptor, al hacer clic, dibuja (o quita) la curva de la mezcla resultante de la última iteración.
 - Esta curva (generalmente una línea amarilla gruesa) representa el mejor ajuste que el modelo encontró, es la conclusión del algoritmo.
- Botón "Ver Historial":
 - Abre una nueva ventana que muestra una tabla detallada con los valores numéricos exactos de (μ_1, σ_1, π_1) y (μ_2, σ_2, π_2) para cada paso de la ejecución.
- Botón "Ver Ecuaciones":
 - Abre una ventana que escribe la fórmula matemática completa de la mezcla (ej. $0.45 \cdot N(x; 12.5, 3.1) + 0.55 \cdot N(y; 18.2, 4.5)$) para cada iteración.
 - Es muy útil para entender visualmente cómo la fórmula cambia a medida que los parámetros se optimizan.
- Botón "Limpiar Historial":
 - Borra los datos de la ventana "Ver Historial".

Exportar Resultados (Guardar Gráfica)

1. Haz clic en el botón "4. Guardar Gráfica".
2. Esto tomará una "foto" (captura de pantalla) de la vista actual del panel de visualización.
3. ¡Importante! Se guardará exactamente lo que ves en ese momento, si tienes activadas "Iter 1" y "Mezcla Final", ambas saldrán en la imagen, o si solo tienes el histograma, solo se guardará eso.
4. El archivo (.png) se guardará en una carpeta llamada em_gallery, que se crea automáticamente en el mismo directorio donde se ejecuta el programa.
5. El nombre del archivo se basa en la variable y el número de iteraciones (ej. radius_nucA_5iters.png).

Módulo: Comparador de Gráficas EM

Esta herramienta te permite cargar y comparar, lado a lado, dos gráficas que hayas guardado previamente desde el módulo EM.

Pasos de uso detallados

Paso 1: Acceder al Comparador

Desde el Menú Principal, haz clic en el botón "Comparador de Gráficas EM".

Paso 2: Cargar la Gráfica Izquierda

1. La ventana se divide en dos secciones: "Gráfica Izquierda" y "Gráfica Derecha".
2. En el panel izquierdo, verás una lista de todos los archivos .png disponibles en la carpeta em_gallery (las imágenes que guardaste anteriormente).
3. Haz clic en el nombre de un archivo en esta lista (ej. PC1_2iters.png).
4. La imagen seleccionada aparecerá en el visor grande de la izquierda.

Paso 3: Cargar la Gráfica Derecha

1. Repite el proceso en el panel derecho, haz clic en otro archivo de la lista derecha (ej. PC1_20iters.png).
2. La imagen seleccionada aparecerá en el visor de la derecha.
3. Ahora puedes comparar visualmente, por ejemplo, cómo se veía el ajuste en la iteración 2 contra cómo se veía en la iteración 20.

Paso 4: Actualizar la Lista (Si es necesario)

1. Si regresaste al módulo EM, guardaste nuevas imágenes y volviste a entrar al comparador, es posible que las nuevas imágenes no aparezcan en la lista.
2. Haz clic en el botón "Actualizar Lista" para forzar al programa a recargar el contenido de la carpeta em_gallery y mostrar los archivos más recientes.

Paso 5: Salir

Haz clic en "Regresar al Menú" para cerrar el comparador y volver a la pantalla principal.