Predicción del número de la lotería Daily 3 de Michigan con Cadenas de Markov

Jhojan Andrés Grisales, Juan Diergo Jimenez, Brayan Acosta, Miguel

**Resumen - Este documento describe la creación de un modelo predictivo para la lotería ‘Daily-3’ de Michigan, EE.UU., utilizando el historial de números ganadores desde el 1 de enero de 2020 hasta la actualidad a dia 17 de Octubre de 2025, utilizando cadenas de Markov y Python. Se desarrolló una aplicación que predice los números de lotería más probables basándose en cada dígito del último número ganador y en el historial de números, permitiendo ingresar cualquier número de tres dígitos para calcular su probabilidad de ganar en cualquier sorteo futuro en el rango de 365 días a partir del día actual. Se realizaron pruebas de aleatoriedad para asegurar la validez de los datos. Con base en los resultados obtenidos del programa, se observó que aunque la predicción del número de ganador de lotería es inherentemente incierta debido a su naturaleza aleatoria, este enfoque ofrece una herramienta básica para analizar el comportamiento de un sistema que funciona de manera estocástica.**

**Palabras Clave: Procesos estocásticos, Cadenas de Markov, Predicción de Lotería, Matriz estocástica, Independencia de datos, Python.**

1. **INTRODUCCIÓN**

Las cadenas de Markov, un concepto fundamental en la teoría de probabilidad, se utilizan para predecir la secuencia de eventos futuros basándose únicamente en el estado actual, ignorando la historia previa.

Este laboratorio aplicó dicha teoría al desarrollo de un modelo predictivo para la lotería "Daily-3 evening" de Michigan, EE. UU., mediante el uso de Python. El propósito fue crear un sistema que prediga los números más probables utilizando resultados históricos desde el 1 de julio de 1999. Se introdujo una funcionalidad para calcular la probabilidad de victoria de cualquier número de tres dígitos propuesto, con base en las cadenas de Markov y matrices de transición de estados. Además, se efectuaron pruebas de correlación de Spearman y chi-cuadrado para asegurar la validez y aleatoriedad de los datos analizados, evaluando si los números ganadores mostraban patrones específicos o si eran completamente aleatorios. Este enfoque no solo demostró la aplicación práctica de las cadenas de Markov en la predicción de eventos aleatorios, sino que también subrayó la importancia de validar los datos mediante pruebas estadísticas rigurosas.

1. **MARCO TEÓRICO**
2. **Cadena de Markov:** Una cadena de Markov es un modelo matemático que describe una secuencia de eventos en la que la probabilidad de que ocurra un evento futuro depende únicamente del evento actual y no de los eventos anteriores. Esta propiedad se conoce como la propiedad de Markov o la propiedad de Markov de primer orden, y es fundamental para la modelización de sistemas estocásticos como los sorteos de lotería.
3. **Matriz de Transición de Estados:** En una cadena de Markov, las probabilidades de transición entre estados se representan mediante una matriz de transición. Cada elemento de la matriz representa la probabilidad de transición de un estado a otro. Esta matriz es esencial para calcular las probabilidades de ocurrencia de estados futuros y para predecir eventos en el sistema modelado.
4. **Prueba de Chi-cuadrado para la aleatoriedad:** La prueba de chi-cuadrado se utiliza para evaluar si los datos siguen una distribución uniforme o si exhiben algún patrón significativo. En este caso, se calcula el estadístico de chi-cuadrado y su correspondiente p-valor para determinar si los números ganadores de la lotería están distribuidos uniformemente o si muestran algún sesgo o patrón.
5. **Prueba de correlación de Spearman:** La correlación de Spearman se utiliza para evaluar la relación monotónica entre dos variables. En este contexto, se utiliza para determinar si hay alguna relación entre los números ganadores de un día y los números ganadores del día siguiente en la lotería.
6. **Predicción de Eventos Futuros:** Utilizando la matriz de transición de estados, podemos predecir la probabilidad de que ocurran ciertos eventos en el futuro. Esto nos permite generar predicciones sobre los números de lotería más probables en los próximos sorteos, basándonos en el historial de resultados pasados.

**Probabilidad de transición de estados:** Es la probabilidad de que el sistema en el futuro cambie de estado *j* dado que estaba en el estado *i* en el tiempo presente, es decir

➔ Probabilidad de transición de estado.

**Probabilidad de estado:** Es la probabilidad de que el sistema se encuentre en un estado determinado en una transición cualquiera (predicción).

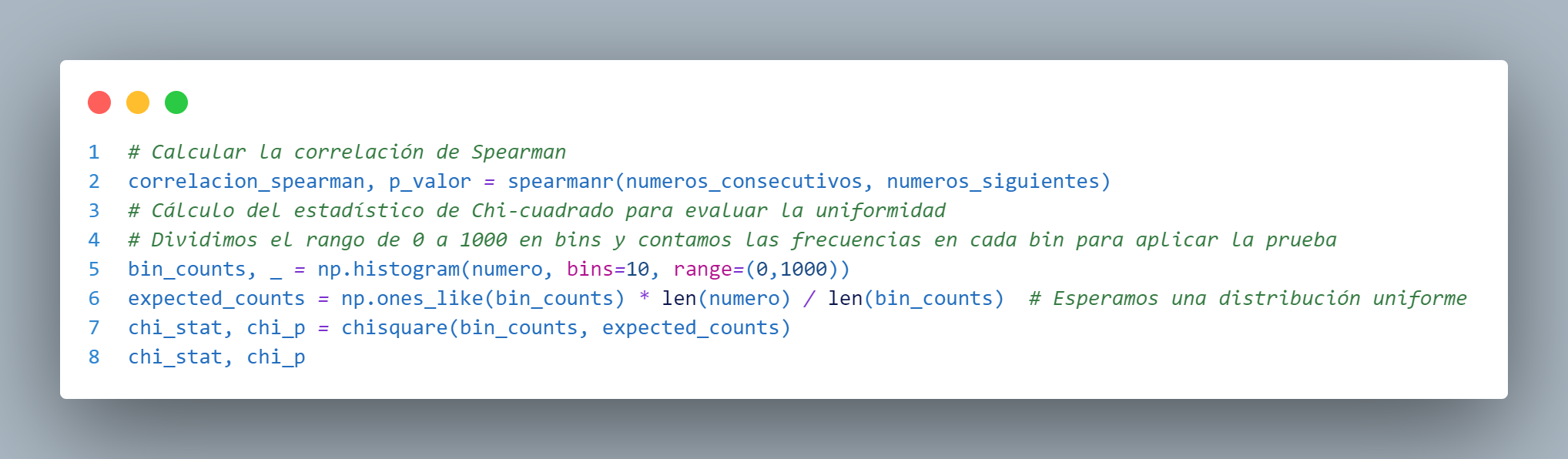
➔ Probabilidad de estado

1. **Ecuaciones de Kolmogorov:** Rigen la dinámica de la cadena de markov
2. ➔ Forma vectorial

➔ Condición inicial

1. **METODOLOGÍA DE DESARROLLO**

Realización de la prueba de Spearman y la prueba de Chi-cuadrado a los números ganadores obtenidos en la lotería "Daily-3 evening" desde el 1 de julio de 1999 hasta la fecha actual. La prueba de Spearman evalúa la correlación entre los números ganadores consecutivos, mientras que la prueba de Chi-cuadrado para la aleatoriedad evalúa si los números ganadores se distribuyen uniformemente o muestran un comportamiento de distribución aleatoria.



**Figura 1.** Fragmento de código Python que verifica la correlación entre los datos mediante la prueba de Spearman y la aleatoriedad mediante la prueba Chi-cuadrado.

**1. Prueba de Spearman:** Se hace uso de la función “spearmanr” de “scipy.stats” para el cálculo del coeficiente de correlación de Spearman y el p-valor asociado. Se ingresan como argumentos los números ganadores consecutivos (“numeros\_consecutivos”) y sus sucesores (“numeros\_siguientes”). El coeficiente de correlación y el p-valor se almacenan en las variables “correlacion\_spearman” y “p\_valor” respectivamente.

**2. Prueba de Chi-cuadrado para la aleatoriedad:** Se realiza con la función “chisquare” de “scipy.stats”, usando las frecuencias observadas obtenidas por “np.histogram” de los números ganadores divididos en 10 bins, y las frecuencias esperadas calculadas como el total de datos dividido por el número de bins. Los argumentos de las frecuencias observadas y esperadas se introducen en “chisquare”, resultando en el cálculo del estadístico de chi-cuadrado y el p-valor, almacenados en “chi\_stat” y “chi\_p” respectivamente.

***¿Los resultados de la loteria son apropiados para ser trabajados en una cadena de markov?***

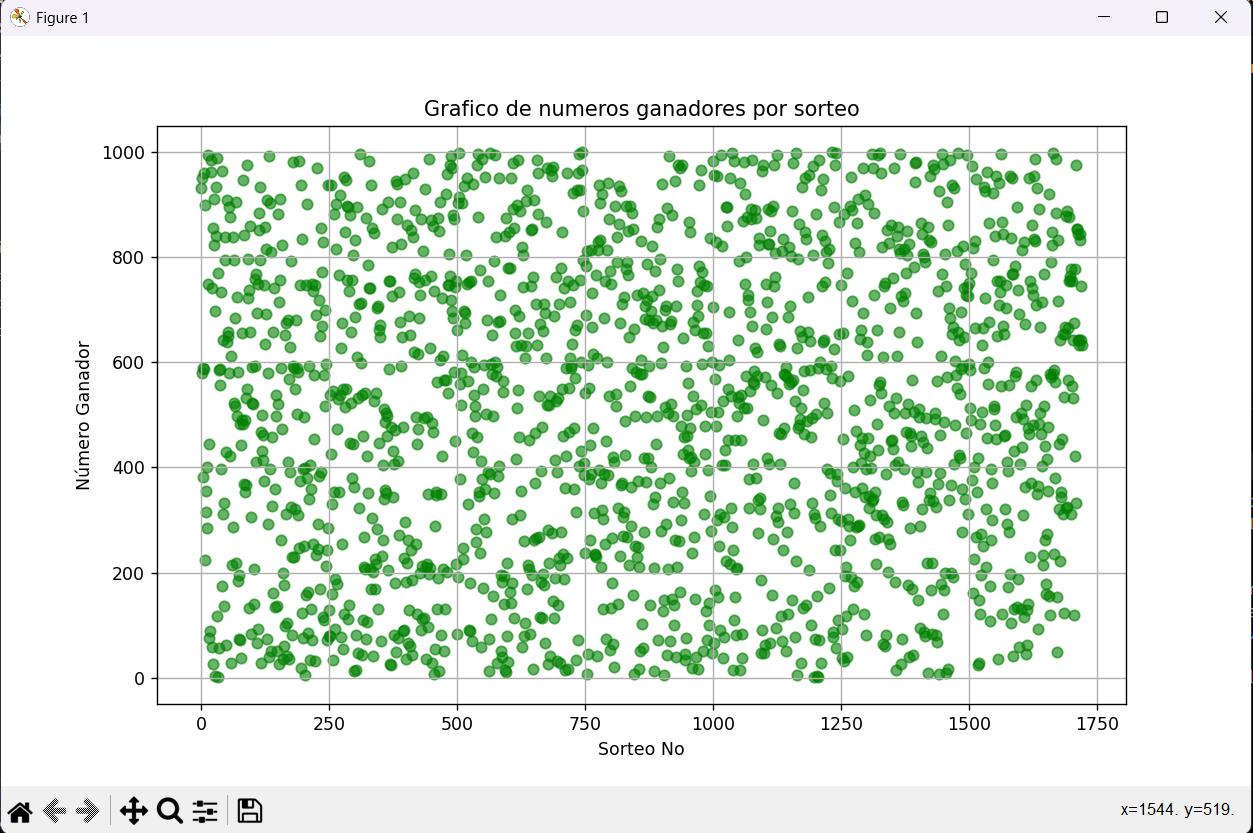
**Prueba de correlación de Spearman:**

El coeficiente de correlación de Spearman obtenido es **-0.001191**, lo que indica una correlación prácticamente nula entre los números ganadores consecutivos. Esto sugiere que *no hay una relación lineal significativa entre los números ganadores en días sucesivos.*

El p-valor asociado a esta prueba es **0.9105**. Dado que este valor es mayor que 0.05, no hay suficiente evidencia para rechazar la hipótesis nula de que no hay correlación entre los números ganadores consecutivos. En otras palabras, *no hay una correlación estadísticamente significativa entre los números ganadores.*

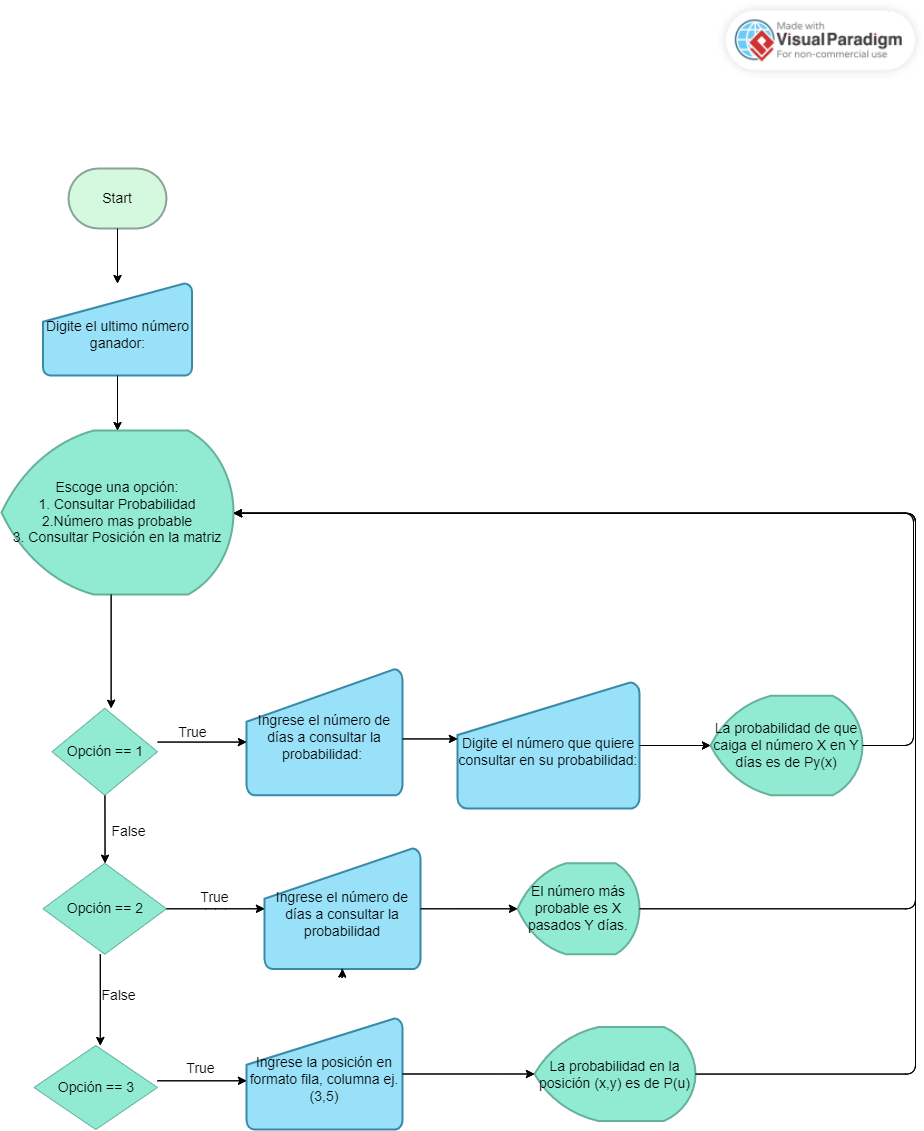
**Prueba de Chi-cuadrado para la Aleatoriedad:**

El estadístico de chi-cuadrado obtenido es 7.0215 y el p-valor asociado a esta prueba es 0.6349. La hipótesis nula de esta prueba es que los datos de la muestra se distribuyen uniformemente o tienen un comportamiento de distribución aleatoria. Dado que el p-valor es mayor que 0.05, no hay suficiente evidencia para rechazar esta hipótesis nula. Esto sugiere que *los números ganadores no muestran un patrón discernible y se comportan de manera aleatoria.*



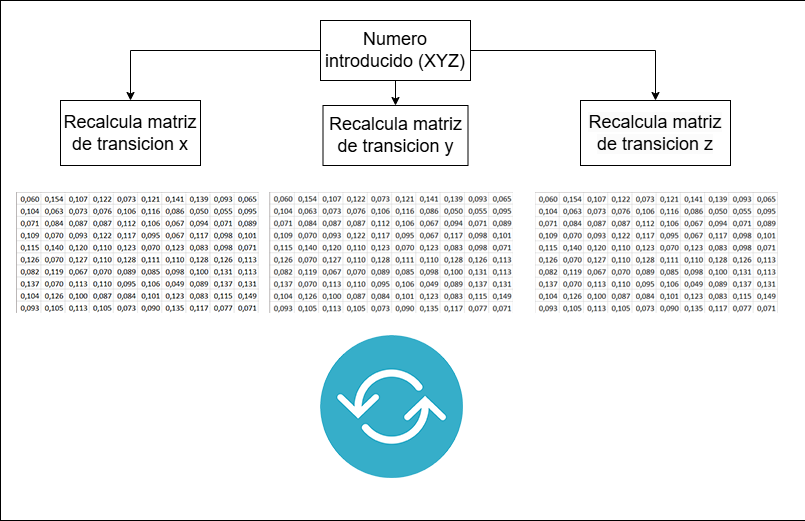
**Figura 2.** Gráfico de dispersión de los números ganadores presentes en la base de datos desde el año 1999 hasta la fecha actual.

El gráfico mostrado en la ***figura 2*** permite visualizar si existe algun patron entre los resultados de loteria de la base de datos utilizada para la simulación, en este caso es posible observar que no existe un patrón visible para inferir que los datos no son aleatorios, por tanto al visualizar este gráfico e interpretar los resultados obtenidos de las pruebas realizadas previamente, se puede afirmar que los datos son totalmente aleatorios y dispersos, y consecuentemente son aptos para utilizarlos en la presente práctica.



**Figura 3.** Diagrama de flujo que describe la lógica del programa que simula la predicción de números de lotería utilizando cadenas de markov.

El diagrama de flujo representa un programa interactivo que inicia pidiendo al usuario ingresar el último número ganador de lotería y luego ofrece tres opciones: calcular la probabilidad de un número específico en un número determinado de días, identificar el número más probable dado un número de días, o consultar la probabilidad de un número en una posición específica de la matriz de transición. Dependiendo de la elección, se solicitan datos adicionales como el número de días para la consulta o la posición en la matriz. Tras presentar los resultados de la opción seleccionada, el programa regresa al menú principal, permitiendo al usuario realizar más consultas o salir del programa.



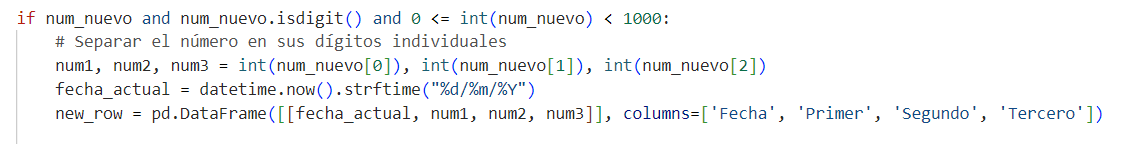
**Figura 4.** Gráfico ilustrativo del funcionamiento del algoritmo con 3 cadenas de Markov.

**Para el desarrollo del código con la simulación utilizando cadenas de Markov, se realizó lo siguiente:**

***¿Cómo se guarda cada número independientemente al digitarlo?***

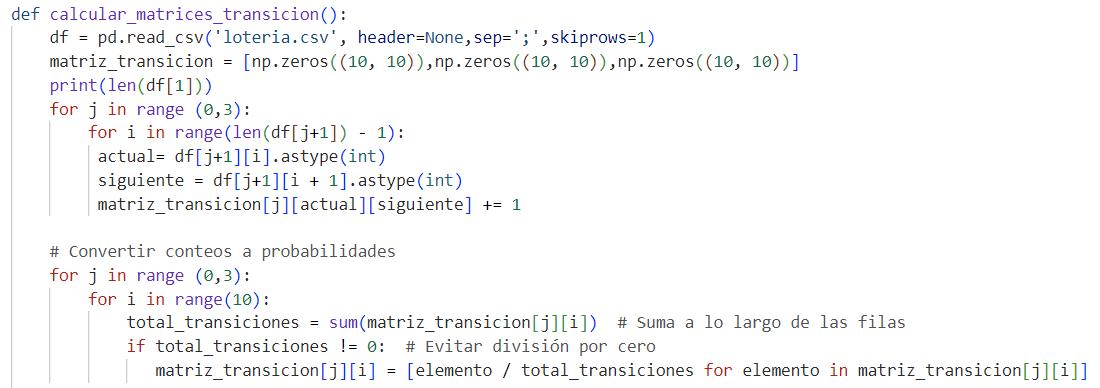
Utilizando la librería Pandas se abre/lee el csv de lotería, que hace el papel de base de datos usada por el algoritmo y divide el número ingresado en sus 3 dígitos individuales para agregarlos en cada columna correspondiente.

La recolección de datos se facilita mediante la función `**añadir\_numero()**`, la cual interactúa con el usuario a través de una GUI desarrollada con Tkinter, solicitando la inserción del último número ganador de la lotería correspondiente.Este dato se agrega automáticamente al histórico de números ganadores, presente en el archivo **‘loteria.csv’**. La validación de entrada asegura la integridad de los datos mediante la verificación de que el número ingresado sea de 3 dígitos (un valor entre 0 y 999).



***¿Cómo se calculan las 3 matrices de transición necesarias para definir las cadena?***

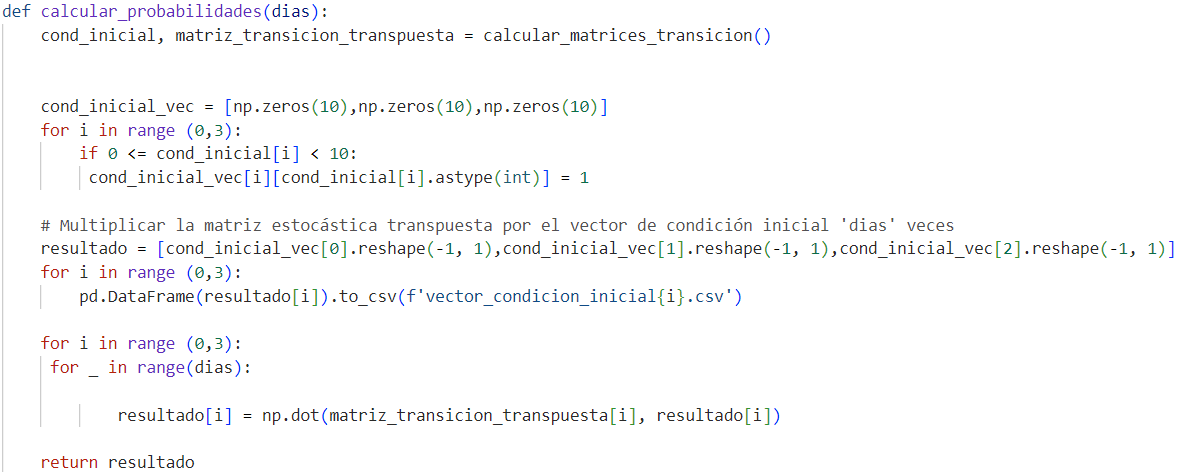
La función **`calcular\_matriz\_transicion()`** es de vital importancia en el análisis estocástico del problema. Utiliza la librería **‘Pandas’** para cargar el dataset histórico desde el archivo CSV y luego emplea **‘Numpy’** para construir 3 matrices de transición de estados. Estas matrices son cuadradas de 10x10, donde cada elemento representa la probabilidad de transición del estado al estado , basada en la frecuencia observada de esta transición en el dataset.



**Figura 5.**  Fragmento de código que corresponde a la función ‘calcular\_matriz\_transicion()’

***¿Cómo se halla la probabilidad de cambio de estado de un número?***

La normalización de filas convierte conteos brutos en probabilidades, transformando la matriz en una representación estocástica válida para análisis de Markov.



**Figura 6.** Fragmento de código que corresponde a lafunción ‘**calcular\_probabilidades(dias)’,** que permite realizar los cálculos de probabilidades solicitadas en el programa, mediante cadenas de Markov.

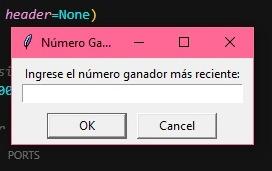
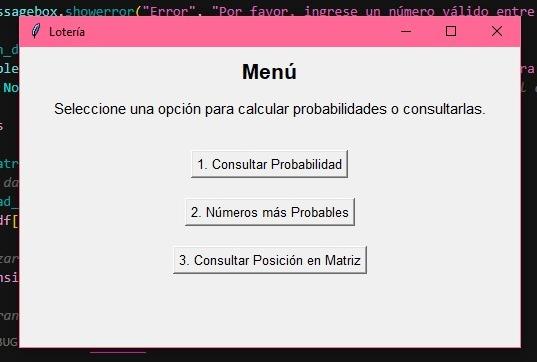
La función en la ***figura 6*** proyecta las probabilidades de ocurrencia de los números en la lotería para días futuros. Inicialmente, asigna a una variable la matriz de transición transpuesta de `**calcular\_matriz\_transicion()**`. Después, crea 3 vectores de condición inicial usando el último número ganador, estableciendo su probabilidad en 1 y las de los demás en 0. Mediante la multiplicación de este vector por la matriz de transición transpuesta a lo largo de los días indicados, simula las transiciones de estado mediante una cadena de Markov, produciendo 3 vectores de probabilidades futuras. Finalmente, confirma que la suma de estas probabilidades es igual a 1 para asegurar la precisión de los cálculos.

La GUI, creada con `**inicializar\_ventana\_principal()**` y funciones relacionadas, ofrece una interfaz interactiva para consultas, adiciones de datos y visualización de predicciones, simplificando la interacción con el análisis estocástico mediante Tkinter.

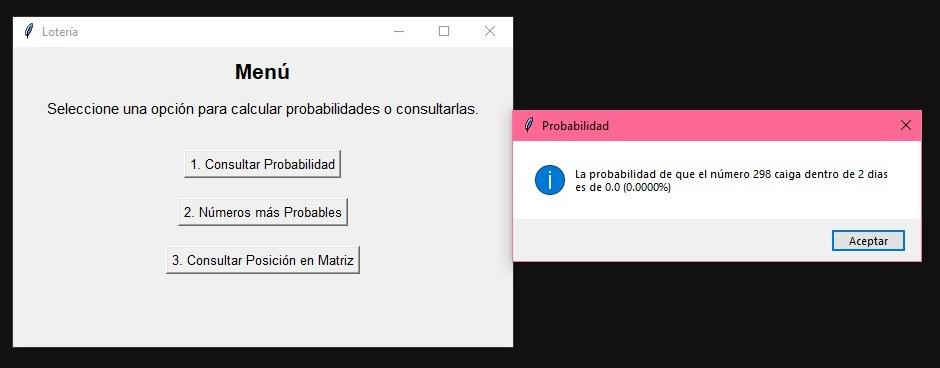
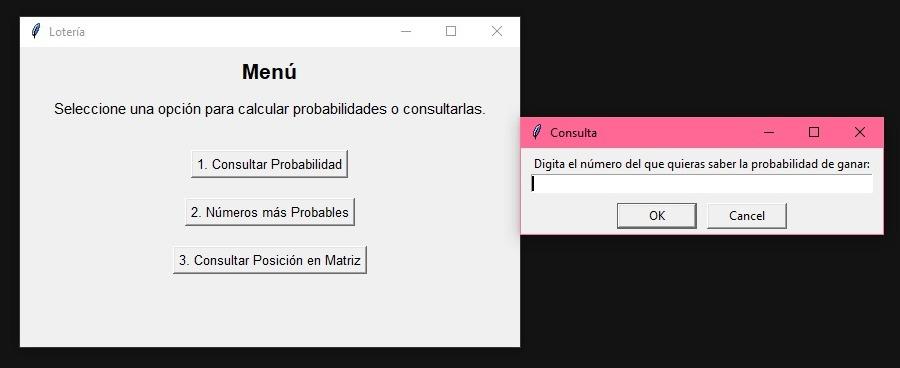
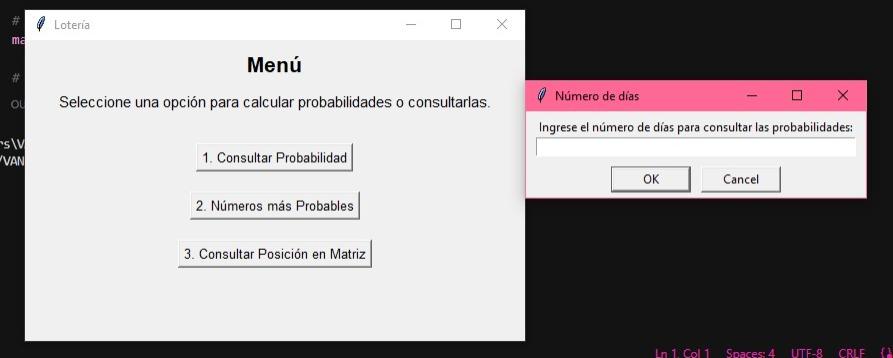
Por otro lado, las funciones `**mostrar\_numero\_mas\_probable()**` y `**consultar\_probabilidad()**` ofrecen al usuario la capacidad de visualizar la probabilidad de números específicos y los más probables de ocurrir, extrayendo información de la matriz de transición y del vector de estado tras los cálculos.

**Validación y Pruebas**

La fase de validación se centra en asegurar la precisión de las matrices de transición y los cálculos de probabilidad. Se realizan pruebas unitarias y de integración para validar cada componente del sistema, desde la carga y manipulación de datos hasta los algoritmos de cálculo de probabilidad y la funcionalidad de la GUI.

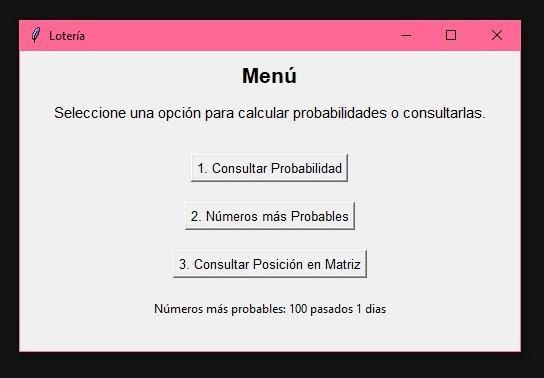
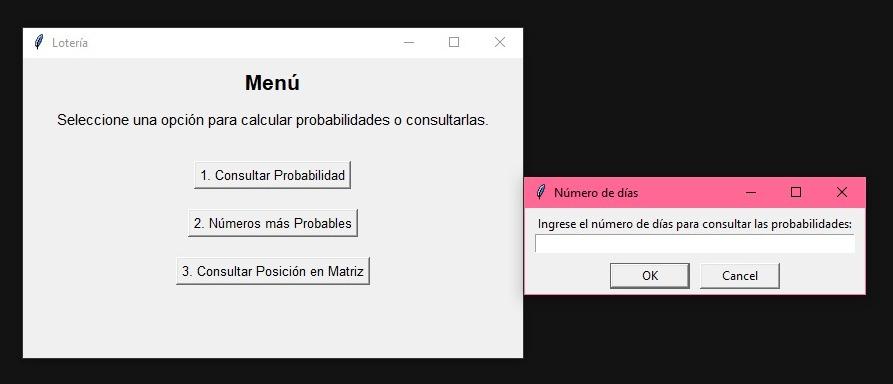
**Figura 7.** Interfaz inicial del programa, campo para ingresar el último número ganador y posteriormente el menú de opciones

La interfaz mostrada en la ***figura 7*** se observa que la entrada inicial del programa es el número ganador más reciente de la lotería, posteriormente al presionar ‘ok’ se abre una ventana con el menú de la aplicación, en el cual se puede ejecutar cualquiera de las tres opciones establecidas.



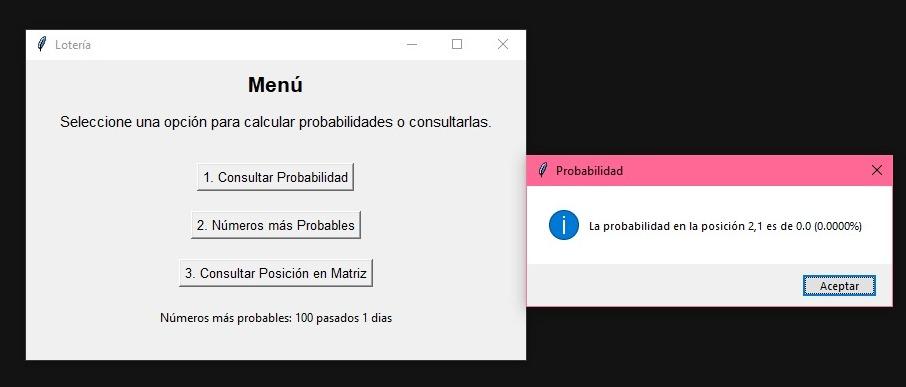
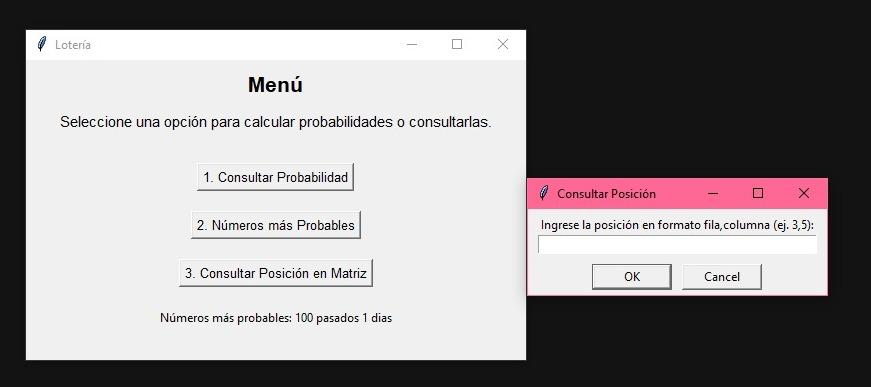
**Figura 8.** Secuencia visual 1 de proceso de cálculo de probabilidad en ‘n’ cantidad de días.

La primera ventana del programa solicita al usuario que introduzca la cantidad de días posteriores para calcular las probabilidades de los números de lotería. Esta entrada define el período para los cálculos de probabilidad futura. En la segunda ventana, el usuario introduce un número de lotería específico y el sistema usa cadenas de Markov para calcular y mostrar la probabilidad de que ese número sea el ganador en el plazo establecido.



**Figura 9.** Secuencia visual 2 de proceso de cálculo de probabilidad en cierto número de días.

La interfaz pide al usuario definir un número de días para evaluar las probabilidades de la lotería, estableciendo así el horizonte temporal para los cálculos. Al ingresar este valor, se determina el período futuro para el cual se desean las probabilidades, basándose en el modelo de cadena de Markov y los datos históricos de la lotería. Este número de días introducidos por el usuario es esencial para guiar la amplitud y el alcance temporal de las predicciones de probabilidad que realizará el sistema.



**Figura 10.** Secuencia visual 3 de (proceso de consulta de la probabilidad presente en un registro específico de la matriz estocástica)

La última opción del programa permite al usuario ingresar un registro específico de la matriz estocástica para consultar la probabilidad presente en este mismo.

1. **CONCLUSIONES**

* A pesar de modelar y predecir los números de lotería utilizando cadenas de Markov, queda claro que la lotería sigue siendo un juego de azar. La incertidumbre inherente a la naturaleza aleatoria de los sorteos significa que no hay garantías absolutas en cuanto a la predicción de los números ganadores.
* Nuestro enfoque proporciona una herramienta básica para analizar el comportamiento de un sistema estocástico. La implementación de cadenas de Markov nos permitió capturar y modelar la dependencia de los resultados en los futuros hasta la tendencia al estado estable que es la probabilidad de un estado cuando las iteraciones de tiempo tienden a un valor muy grande.
* El cálculo de probabilidades de una lotería tipo Pick 3 usando cadenas de Markov, requiere un histórico de datos extenso, lo cual se puede ver reflejado en que el set de datos de la figura 3 usados correspondientes a un aproximado de 24 años de muestreo no abarca un porcentaje significativo de transiciones de la matriz estocástica.
* Este laboratorio demostró la utilidad de aplicar conceptos teóricos, como las cadenas de Markov, en la resolución de problemas prácticos. La combinación de teoría y práctica nos permitió desarrollar una herramienta funcional que puede tener aplicaciones en la predicción de otros fenómenos estocásticos.
* La ausencia de correlación estadísticamente significativa entre los números ganadores de la lotería, indica que el proceso de selección de números es aleatorio y que cada sorteo es independiente de los anteriores.

1. **REFERENCIAS**

* Jones, R. H. (2016). Introduction to Stochastic Processes with R. Springer.
* Taylor, H. M., & Karlin, S. (1998). An introduction to stochastic modeling (Vol. 139). Academic press.