Guía de ejercicios 1

Probabilidad y generación de números aleatorios

▷ **Ejercicio 1.**
Determine el espacio muestral de los siguientes experimentos aleatorios:

- Arrojar una moneda
- Arrojar un dado de seis lados
- Arrojar dos monedas iguales al mismo tiempo
- Arrojar tres monedas diferentes al mismo tiempo
- Arrojar dos monedas iguales en orden

▷ **Ejercicio 2.**
Sea un experimento aleatorio con eventos A y B en el espacio muestral. Decir si es verdadero o falso, justificando la respuesta:

- $\mathrm{P(B - A) = P(B) - P(A)}$
- $\mathrm{P(B - A) = P(B) - P(A \cap B)}$
- $\mathrm{P(A) = 0} \implies \mathrm{P(A \cap B) = 0}$
- $\mathrm{P(A \cup B) = 0} \implies \mathrm{P(A) = P(B) = P(A \cap B) = 0}$

▷ **Ejercicio 3.**
Verifique que al lanzar dos dados la probabilidad de obtener 7 u 11 es 2/9.

▷ **Ejercicio 4.**
Se lanzan dos monedas. A es el evento “dos caras” y B es el evento “dos secas”. Calcule la probabilidad de A $\vee$ B. ¿Son A y B mutuamente excluyentes? ¿Son complementarios?

▷ **Ejercicio 5.**
En una población de ratones, el 40% de la población tiene el pelo negro, el 25% tiene ojos rojos y el 15% tiene el pelo negro y ojos rojos. Se escoge un ratón al azar:

- Si tiene el pelo negro, ¿ Cuál es la probabilidad que tambien tenga ojos rojos?
- Si tiene ojos rojos, ¿ Cúal es la probabilidad de que no tenga pelo negro?

▷ **Ejercicio 6.**
Se conoce que en una colmena existen aproximadamente 100 zánganos y 500 obreras. Si se eligen 3 abejas sin reposición, hallar la probabilidad de que:

- todas sean obreras,
- las 2 primeras sean zánganos y la tercera sea obrera,
- por lo menos una sea obrera.

▷ **Ejercicio 7.**

En la tabla siguiente se presentan los resultados del examen de 1000 individuos en relación al sexo y el uso de lentes. Utilícela para determinar

- ¿cúal es la probabilidad de que un individuo sea varón dado que usa lentes?,
- ¿cúal es la probabilidad de que un individuo use lentes dado que es mujer?,
- ¿es el uso de lentes independiente del sexo?

	Sin lentes	Con lentes	Total
Varón	440	160	600
Mujer	160	240	400
Total	600	400	1000

▷ **Ejercicio 8.**

Una caja contiene 6 cubos rojos y 4 verdes y una segunda caja contiene 7 cubos rojos y 3 verdes. Se elige al azar un cubo de la primera caja y se pone en la segunda caja. Luego se selecciona al azar un cubo de la segunda caja y se pone en la primera caja.

- ¿Cuál es la probabilidad de que se seleccione un cubo rojo de la primera caja y un cubo rojo de la segunda caja?
- Al finalizar el proceso de selección ¿cuál es la probabilidad de que los números de cubos rojos y verdes de la primera caja sean idénticos a los que había al comienzo?

▷ **Ejercicio 9.**

Juan debe tomar un test para conocer si padece cierta enfermedad. Denotamos el estado de salud de Juan por la variable a y el resultado del test por b. Si $a = 1$, Juan tiene la enfermedad, y si $a = 0$ Juan no tiene la enfermedad. El resultado del test es "positivo" (b=1) o "negativo" (b=0). El test es 95% confiable, esto es, en el 95% de los casos de gente que padece la enfermedad, el test retorna un resultado positivo, y en el 95% de los casos de personas que no padecen la enfermedad, un resultado negativo es obtenido.

Se sabe además que el 1% de la personas de la edad de Juan padecen la enfermedad. Si Juan se realiza el test y el resultado es positivo, ¿Cúal es la probabilidad de que Juan tenga la enfermedad?

▷ **Ejercicio 10.**

- Se arrojan dos dados comunes. ¿Cuál es la distribución de probabilidad de la suma de los valores de ambos? ¿Cuál es la distribución de probabilidad del valor absoluto de la diferencia de ambos?.
- Ahora 100 dados comunes son arrojados. ¿Cuál es aproximadamente la distribución de probabilidad de la suma de los valores?. Estime su media y desviación estándar.
- ¿Cómo podrán ser numerados dos dados (con los números del 1 al 6) de modo que cuando arrojamos dos dados la suma de los valores tenga una distribución de probabilidad uniforme en los enteros del 1 al 12?
- ¿Existe alguna manera de numerar 100 dados con enteros tal que la distribución de probabilidad de la suma sea uniforme?

▷ **Ejercicio 11.**

Compute valores de expectación y varianzas para las distribuciones mencionadas en la clase sobre *Distribuciones de probabilidad*

▷ **Ejercicio 12.**

Supongamos que observamos una fuente radioactiva que emite partículas a una tasa descripta por una distribución exponencial con tiempo característico $\tau = 1$. Encuentre la probabilidad de que una partícula (no necesariamente la primera) aparezca

(a) dentro del siguiente segundo,

(b) dentro de los siguientes tres segundos,

(c) dentro del tercer al cuarto segundo desde ahora

(d) después de cuatro segundos desde ahora.

▷ **Ejercicio 13.**

Se tiene un cristal con dos tipos de impurezas, una que absorbe fotones de una frecuencia particular sin liberar un electrón y otra que libera un electrón cuando absorbe un fotón. El cristal tiene el mismo número de impurezas de cada clase, pero la sección eficaz de la primera es 99 veces mayor que la de la segunda. El cristal es lo suficientemente grueso de modo que todos los fotones son absorbidos. ¿Cuál es la probabilidad de que al menos tres electrones sean liberados?

▷ **Ejercicio 14.**

La distribución binomial se aproxima a la distribución normal si el número de realizaciones es grande. Compare el cálculo exacto de probabilidad de obtener 20 veces cara en 40 lanzamientos de una moneda, con la aproximación obtenida con la distribución normal.

▷ **Ejercicio 15.**

Los clientes ingresan a un salón de ventas según un proceso de Poisson homogéneo, con una tasa de 25 clientes por hora.

(a) ¿Cuál es la probabilidad de que entren 10 clientes en la primera hora?
(b) ¿Cuál es la probabilidad de que entren por lo menos 4 clientes en la primera hora?
(c) ¿Cuál es la probabilidad de que hayan entrado 12 clientes en la segunda hora si entraron 30 en las primeras dos horas?

▷ **Ejercicio 16.** *Cadenas de Markov y Predicción del Clima*

Considera un modelo simple de predicción del clima donde la probabilidad de que un día soleado siga a otro día soleado es del 70%, y la probabilidad de que un día lluvioso siga a un día soleado es del 30%. La probabilidad de que un día lluvioso siga a otro día lluvioso es del 60%, y la probabilidad de que un día soleado siga a un día lluvioso es del 40%.

(a) Construye la matriz de transición para esta cadena de Markov.
(b) Si hoy es un día soleado, ¿cuál es la probabilidad de que sea soleado dentro de dos días?

▷ **Ejercicio 17.** *Teorema Central del Límite en Acción*

Simula el lanzamiento de un dado de seis caras 10, 100 y 1000 veces, y calcula el valor promedio obtenido en cada caso. Traza la distribución de los promedios para 1000 simulaciones y compárala con una distribución normal.

- Explica cómo se aplica el Teorema Central del Límite a tus resultados.

▶ **Ejercicio 18.** *Generación de números aleatorios uniformes*

(a) Escriba un programa para generar números aleatorios en el rango $[0, 1]$ usando el método de congruencia lineal. Intente utilizando la siguiente elección de parámetros $(a, c, M, x1) = (57, 1, 256, 10)$. Determine el período, esto es, cuántos números son generados antes de que la secuencia se repita. Tome la secuencia de números generada y busque correlaciones, observando pares de números sucesivos (no los conecte con líneas). ¿Puede esta secuencia usarse para un trabajo serio?. Testee el método con constantes razonables como las estudiadas. Evalúe el momento de orden k (k=1, 3 y 7) de la distribución de (N=10, 100 y 1000) números obtenida, y compare con $\frac{1}{k+1}$.
(b) Compute un número random Δx en el rango $[-\sqrt{2}, \sqrt{2}]$ y otro Δy en el mismo rango y realize varias (K=10) caminatas aleatorias de $N = 1000$ pasos, comenzando en el origen del plano (cada una con diferentes semillas). Calcule el valor de expectación de la distancia al origen R como función del paso N (y de $\sqrt{N}$).

▶ **Ejercicio 19.** *Generador de Fibonacci con Retardo*

El generador de Fibonacci con retardo (lagged Fibonacci generator) es un algoritmo para generar números pseudoaleatorios basado en la recurrencia de Fibonacci generalizada.

Propuesto por George Marsaglia (1980s), es predecesor de generadores más modernos como Mersenne Twister. Se basa en una relación de recurrencia que combina números anteriores de la secuencia con retardos específicos, lo que permite generar números aleatorios de alta calidad y largo período.

La relación de recurrencia general es:

$$X_n = (X_{n-j} \star X_{n-k}) \mod m$$

donde:

- X_n es el n-ésimo número de la secuencia
- j y k son retardos (lags) con $j < k$
- $\star$ es una operación binaria (+, -, $\times$, $\oplus$, etc.)
- m es el módulo (usualmente 2^{32})

Algunas de sus propiedades son:

- Periodo largo: Con parámetros adecuados, el período puede ser $(2^k - 1) \cdot 2^{m-1}$
- Calidad: Mejor que generadores congruenciales lineales simples
- Eficiencia: Computacionalmente eficiente con operaciones simples

Los parámetros j, k y m son cruciales para la calidad del generador. Un ejemplo común es el generador de Fibonacci con retardo con los siguientes parámetros:

- $j = 24$, $k = 55$, $m = 2^{32}$
- Operación: suma o resta
- Semilla inicial: array de k valores

(a) Implemente en Python un generador de Fibonacci con retardo
(b) Use los parámetros: $j = 24$, $k = 55$, $m = 2^{32}$
(c) La operación debe ser suma (+)
(d) Inicialice el array de estado con un generador congruencial lineal
(e) Genere 10,000 números aleatorios en el rango $[0, 1)$

(f) Realice las siguientes pruebas:
   (1) Calcule la media y varianza muestral. Compare con los valores teóricos esperados.
   (2) Genere un histograma y verifique visualmente la uniformidad

(g) Compare los resultados con `numpy.random.random()`

▶ **Ejercicio 20.** *Coeficiente de Correlación de Pearson*

Denotado por r, es una medida estadística que cuantifica el grado de relación lineal entre dos variables cuantitativas. Fue desarrollado por Karl Pearson en 1895 y es ampliamente utilizado en diversas disciplinas científicas.

Se define como:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

donde:

- n es el número de observaciones
- x_i e y_i son los valores individuales de las variables
- $\bar{x}$ e $\bar{y}$ son las medias muestrales de x e y respectivamente

El siguiente código implementa el cálculo del coeficiente de correlación de Pearson en Python. Calcule el coeficiente de correlación de Pearson, considerando retardos de 1, 2, 3, 5, 7 y 10 saltos, para los dos generadores de números aleatorios.

```python

def pearson_correlation(x, y):
    """
    Calcula el coeficiente de correlacion de Pearson entre dos arrays

    Parameters:
    x, y: arrays de igual longitud

    Returns:
    r: coeficiente de correlacion de Pearson
    """
    # Verificar que tienen la misma longitud
    if len(x) != len(y):
        raise ValueError("Los arrays deben tener la misma longitud")

    n = len(x)

    # Calcular medias
    mean_x = np.mean(x)
    mean_y = np.mean(y)

    # Calcular numerador y denominador
    numerator = np.sum((x - mean_x) * (y - mean_y))
    denominator = np.sqrt(np.sum((x - mean_x)**2) * np.sum((y - mean_y)**2))

    # Evitar division por cero
    if denominator == 0:
        return 0

    return numerator / denominator.
```

Listing 1: Implementación del Generador de Fibonacci

▶ **Ejercicio 21.** *Monty Hall*

En un programa de TV de juegos, a un participante se le muestran tres puertas y se le pide que elija una. Detrás de una de las puertas hay un auto (el premio mayor) mientras que detrás de las otras dos hay una cabra. Luego de que el participante elije su puerta, el presentador, quien sabe qué hay detrás de cada una de las puertas, abre una de las dos puertas restantes la cual esconde una cabra. Se le pregunta entonces al participante si desea cambiar su elección o mantener la misma.

(a) Cuál es la probabilidad de ganar el auto si el participante cambia su elección?
(b) Cuál es la probabilidad de ganar el auto si el participate no cambia su elección?

▶ **Ejercicio 22.**

En un catálogo de galaxias hay 4 tipos de galaxias, con probabilidades 0.4, 0.3, 0.2, y 0.1 para los tipos "elíptica", "espiral", "enana" e "irregular", respectivamente. Escriba un programa de PYTHON que genere aleatoriamente tipos de galaxias de tal forma que sigan la distribución dada.

▶ **Ejercicio 23.**

Un experimento aleatorio consiste en arrojar dos dados equilibrados y anotar la suma de los números resultantes en cada uno.

(a) Identifique claramente cuál es el espacio muestral y la variable aleatoria.
(b) Asumiendo que los 6 posibles resultados de cada dado son equiprobables, obtenga la distribución de probabilidad de la variable aleatoria.
(c) Genere un conjunto de valores de la variable aleatoria, utilizando un generador de números aleatorios, a partir de la distribución de probabilidad teórica. Compare la distribución empírica resultante con la función densidad de probabilidad teórica.
(d) Realice una simulación del experimento aleatorio, es decir, pares de valores correspondientes a los dos dados, y calcule la variable aleatoria "suma". Obtenga la distribución de probabilidad empírica de la suma y compare con la distribución teórica.

Deberá entregar un informe sobre los ejercicios marcados con "▶". En dicho informe deberá incluir la resolución de los ejercicios, junto con los códigos empleados y los gráficos correspondientes, acompañados de una breve introducción y conclusiones.

Fecha límite de entrega de informes: viernes 05 de septiembre, a las 13:00 horas.