

# Fenómenos aleatorios e incertidumbre

Carlos E. Rodríguez

Probabilidad aplicada  
Licenciatura en Ciencia de Datos

Julio 2023



# Contenido

- ① Motivación
- ② Ejemplo 1
- ③ Ejemplo 2
- ④ En resumen
- ⑤ Probabilidad

## 1. El número de

- Productos defectuosos que produce una fábrica.
- Productos que regresamos a Amazon.
- Accidentes en cierta avenida.
- Robos y homicidios en una cierta alcaldía o colonia.
- Personas que entran al metro en una cierta estación.

Son fenómenos que parecieran no seguir ninguna ley. Otros más sutiles como

2. La manera (o patrón) en la que los usuarios exploran un sitio de internet.
3. La forma en la que navegamos por FB, Instagram, twitter, Netflix, etc.
4. entre muchos, pero muchos otros casos...

son fenómenos que suceden todo el tiempo a nuestro alrededor y que parecieran ser “aleatorios”.

# Fenómenos aleatorios: Ejemplo 1

Los ingenieros, matemáticos y físicos están bien equipados con herramientas para resolver problemas que caen en el mundo determinista. Pero qué haríamos si se nos presenta el siguiente problema

No	Candidato A	Candidato B
1	309	317
2	298	285
3	313	307
4	316	302
5	295	296
6	290	342
7	321	324
8	318	328
9	290	327
10	291	329

Dos candidatos compiten por un puesto de elección popular y sólo se cuenta con una muestra de votos. ¿Qué candidato ganó la elección?

# Fenómenos aleatorios: Ejemplo 1, intuición

Para cualquier persona es fácil determinar cuál de entre dos números es el más grande, pero en este caso

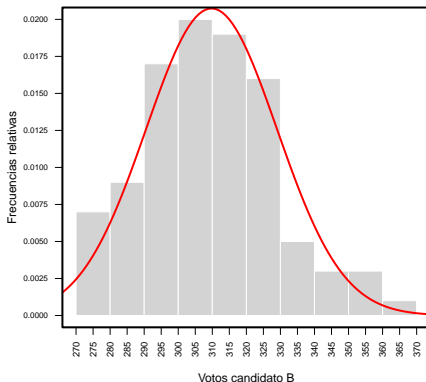
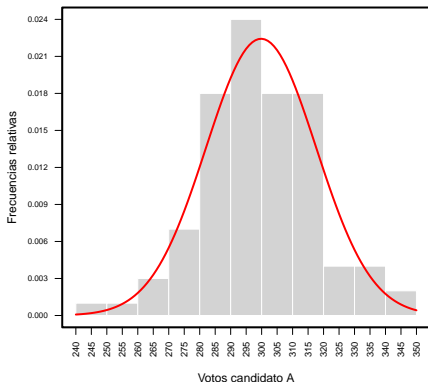
- El problema es la “varianza aleatoria” entre los votos de cada casilla.
- Intuitivamente, ¿qué haríamos?
  1. Calculamos promedios (¿por qué?): el promedio de votos en favor del candidato A es 304.1 y el promedio de votos en favor del candidato B es 315.7.
  2. Con esto, ¿es posible decir que ganó el Candidato B?
  3. ¿Qué pasaría si obtuviéramos otra muestra de casillas?
  4. ¿Cómo se obtuvo la primer muestra?, ¿por qué?
  5. ¿Obtendríamos las mismas casillas en muestra?
  6. ¿Los promedios serían los mismos?
  7. ¿Es buena idea usar los promedios?

¿Cómo podemos introducir una estructura matemática formal en el problema anterior?

- Diremos que existe un modelo  $X$  que describe el comportamiento -agregado- del número de votos del Candidato A en cada casilla.
- Este  $X$  es un objeto que puede tomar distintos valores de manera aleatoria dentro de un cierto conjunto.
- Los votos en favor del Candidato A en la muestra de casillas, i.e.  $x_1 = 309, x_2 = 298, \dots, x_{10} = 291$  son algunos de los valores que puede tomar  $X$ .
- Es imposible (o sumamente difícil) predecir el valor exacto que va a tomar  $X$ . En otras palabras  $\rightarrow$  ¿ $x_{11}$ ?
- Sin embargo, analizando el comportamiento -agregado- de los valores que toma  $X$  podremos identificar una ley (o patrón) que rige el comportamiento de  $X$ .

# Fenómenos aleatorios: Ejemplo 1, aplicación

Para tener una mejor visión de las cosas, supongamos que se tienen muestras de votos en 100 casillas y generamos una visualización del comportamiento agregado de los votos para cada candidato.



## Fenómenos aleatorios: Ejemplo 2

Una medida de la calidad en el proceso de fabricación de láminas de vidrio es el número de defectos en cada lámina (pequeñas partículas “incluidas” en la lámina). Antes de entregar un lote de 1,000 láminas a sus clientes, un fabricante envía una muestra de 60 láminas al laboratorio de Control de Calidad (CC) de la empresa. Se utiliza un dispositivo de escaneo óptico para determinar el número de defectos en cada lámina. Se observó que

- 22, 0 defectos.
- 23, 1 defecto.
- 11, 2 defectos.
- 1, 3 defectos.
- 2, 4 defectos.
- 1, 5 defectos.



## Fenómenos aleatorios: Ejemplo 2

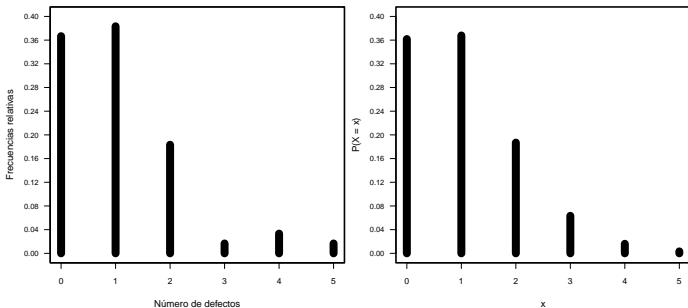
El encargado de CC de la empresa sabe lo siguiente:

1. Sólo se analizó una muestra de 60 láminas de un total de 1,000.
2. Láminas con 3 o menos defectos son aceptables.
3. Láminas con 4 defectos son aceptables siempre que un lote de 1,000 láminas no contenga más de 20 láminas de este tipo.
4. Una lámina con 5 o más defectos es inaceptable.
5. El proceso de fabricación fue diseñado de manera que, cuando se sigue correctamente, no debería haber más de 3 láminas con 5 o más defectos en un lote de 1,000 láminas.

¿Qué es lo que tiene que determinar el encargado de CC de la empresa?

De nuevo asumimos que existe un modelo  $X$  que modela el comportamiento agregado del número de defecto en cada lámina.

1. Necesitaríamos calcular las siguientes probabilidades  $P(X = x)$ , para  $x = 0, 1, 2, \dots$ ,
2. Aproximadamente habrá  $1,000P(X \leq 3)$  láminas con 3 defectos o menos.
3. Aproximadamente habrá  $1,000P(X = 4)$  láminas con 4 defectos.
4. Aproximadamente habrá  $1,000P(X \geq 5)$  láminas con 5 defectos o más.



1. A nuestro alrededor suceden todo el tiempo cualquier cantidad de fenómenos aleatorios.
2. En estos casos la variabilidad aleatoria es el reto a vencer.
3. Analizando el comportamiento agregado de estos fenómenos es posible encontrar patrones o tendencias que nos ayuden a entender lo que sucede.
4. El histograma y la gráfica de barras son dos herramientas para visualizar el comportamiento agregado de fenómenos aleatorios.
5. Existen modelos matemáticos capaces de describir estos comportamientos aleatorios. Por ejemplo:

$$\text{Ejemplo 1: Normal: } f(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2},$$

$$\text{Ejemplo 2: Poisson: } f(x|\theta) = \frac{e^{-\theta}\theta^x}{x!}.$$

Existen principalmente dos fuentes de aleatoriedad:

1. **Intrínseca.** Es parte natural del fenómeno bajo estudio y la incertidumbre que produce no se puede eliminar. Por ejemplo: el comportamiento de cualquier población (animales, insectos, seres humanos, etc), urnas, dados, cartas, etc.
2. **Subjetiva.** Se debe al conocimiento parcial (ignorancia) de las personas. La incertidumbre que produce puede disminuirse o incluso eliminarse. Por ejemplo: errores de medición, producida por un modelo, información en muestras de poblaciones finitas, etc.

# ¿Cómo podemos enfrentar la incertidumbre que plantean los fenómenos aleatorios?

1. Analizar observaciones/eventos de manera individual no es buena idea.
2. Analizar el comportamiento agregado de muchas observaciones/eventos en busca de patrones es la mejor estrategia.
3. Determinar todos los posibles valores/eventos que pueden tomar las observaciones.
4. Asignar probabilidades a cada uno de estos valores/eventos.
5. Tomar decisiones o diseñar estrategias con base en estas probabilidades.

# ¿Qué es una probabilidad?

Es un número entre 0 y 1 que indica el grado de incertidumbre que se tiene de que observemos cierto valor/evento.

- Si un valor/evento tiene probabilidad 1, entonces tenemos la certeza de que SÍ sucederá.
- Si un valor/evento tiene probabilidad 0, entonces tenemos la certeza de que NO sucederá.
- La máxima incertidumbre se da cuando la probabilidad es 0.5.

# ¿Cómo calculamos la probabilidad de que un evento $A$ ocurra?

Existen varias maneras para obtener probabilidades.

- Exacta bajo el supuesto de eventos igualmente probables y usando combinatoria.

$$P(A) = \frac{N_A}{N},$$

- Asintótica bajo los supuestos de que es posible repetir un experimento un número infinito de veces y bajo las mismas condiciones cada vez.

$$P(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n},$$

- Subjetiva. Una medida de la creencia personal que alguien tiene de que el evento  $A$  sucederá,  $P(A)$ .