

ESTADISTICA COMPUTACIONAL

Programación dinámica: **Técnica comúnmente usada para la optimización de ciertos problemas con el fin de reducir su tiempo de ejecución**; se usa especialmente en aquellos casos en los que se puede hallar una solución óptima, en problemas cada vez más pequeños (Subestructura optima) o en donde repetimos el mismo computo una y otra vez, Fibonacci (problemas empalmados).

Para optimizar estos problemas se usa la memorización (memoization), guardamos el resultado de los cálculos en una estructura de datos de fácil acceso (diccionario) y accedemos a estos resultados en lugar de hacer el computo. **Intercambio de tiempo por espacio.**

Caminos aleatorios: Elegir dentro de un set de decisiones correctas o comportamientos de manera aleatoria, dado un input es probable que tengamos un output diferente.

Donde el objeto se mueve de manera aleatoria se le llama campo.

Programación estocástica: Permite introducir la **aleatoriedad** a los programas, se usa en casos en donde las distribuciones de probabilidad de un problema se conocen o pueden ser estimadas, esta trabaja a su vez con inferencia estadística.

- Un programa es **determinístico** si con el mismo input recibimos el mismo output.

Calculo de probabilidades: Medida de la certidumbre de que algo suceda o no, calculamos la probabilidad entre 0 (no va a suceder) – 1 (seguro que suceda)

Leyes:

- Ley del complemento: La probabilidad de que algo suceda + la probabilidad de que ese algo no suceda tiene que ser igual a 1.

$$P(A) + P(\neg A) = 1$$

- Ley multiplicativa: Determina cual es la probabilidad de que un evento suceda y otro evento suceda, siempre y cuando los dos eventos sean “independientes”

$P(A \text{ y } B) = P(A) * P(B)$ ¡La probabilidad de ambas cosas siempre es menor que la de 1 sola cosa!

- Ley aditiva: Cual es la probabilidad de que A suceda o B suceda

$$P(A \text{ o } B) = P(A) + P(B) \text{ (eventos exclusivos)}$$

$$P(A \text{ o } B) = P(A) + P(B) - P(A \text{ y } B) \text{ (eventos no exclusivos)}$$

Inferencia estadística: Ayuda al no saber las probabilidades de eventos simples, sus técnicas nos permiten inferir/concluir propiedades de una población a partir de una muestra aleatoria.

Entre mas pruebas de una pequeña población la media de esta se acercará cada vez más a la media de la población total debido a la **ley de los grandes números**.

Nos provee rangos de validez e intervalos de confianza para poder calcular la media de una población a partir de muestreos.

Ley de los grandes números: entre más cantidad de muestras y de pruebas más se acercará la desviación del resultado a 0.

Falacia del apostador: Después de evento extremo ocurrirán eventos menos extremos para nivelar la media.

Regresión a la media: Después de un evento extremo, el siguiente evento probablemente será menos extremo.

Media: Promedio de los valores, permite inferir en donde se encuentran la mayoría de los valores de la población objetivo. La media de una población (μ – miu) y la media de una muestra (\bar{X} - x bar)

Varianza y desviación estándar: Medidas de propagación que tan lejos o cerca a la media están los valores.

Varianza: Mide que tan propagados de la media se encuentra un conjunto de valores, nos indica que tan dispersos están los datos.

Desviación estándar: Raíz cuadrada de la varianza, habla en las mismas unidades que la media.

Media = 9.85

Desviación estándar = 5.5

Podemos esperar que la mayoría de los datos se encuentren entre 15 y 4

Distribución normal: Se define por la media y por la desviación estándar (con estos dos se puede describir cualquier distribución normal), cuáles son las probabilidades de algún suceso.

Se puede convertir una distribución que no es normal en una normal a través del teorema de limite central.

Teorema del limite central: Permite convertir cualquier distribución en una distribución normal. Se saca la media de c/u de las muestras y mientras más grande sea el tamaño de las muestras, las medias de estas muestras tienden cada vez más a una distribución normal.

Regla empírica: Calcular como se distribuyen dentro de una distribución normal los datos.

Muestreo: Importante cuando no hay capacidad computacional para calcular todo, coger muestras para obtener resultados validos estadísticamente.

Muestreo probabilístico:

- Muestreo aleatorio: Se hace cuando la población posee características similares.
- Muestreo estratificado: Genera subgrupos relevantes para no sesgar la muestra cuando existen diferencias relevantes en la población.

Regresión lineal: Aproximar un conjunto de datos a una función no necesariamente lineal.