Reinforcement Learning

El '**RL**' es un método de '**Aprendizaje Automático**' para que un '**Agente**' guíe su propio aprendizaje por medio de un esquema de '**castigo y recompensa**', incentivando al agente maximizar la '**recompensa**' (numérica).

Se Diferencian

Reinforcement Learning

Se busca maximizar la recompensa.

Supervised Learning And Unsupervised Learning

Se busca reducir el error.

Un modelo de '**Reinforcement Learning'** esta conformado por 6 variables

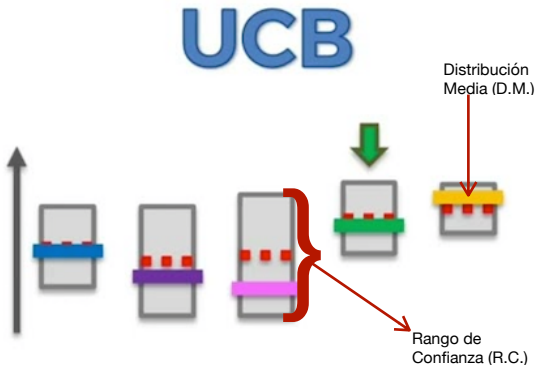
- **Ambiente**
  - Entorno con reglas y limitaciones establecidas donde el '**Agente**' interactúa.
- **Estado**
  - Situación/Indicador actual del '**Ambiente**' en un momento determinado.
- **Agente**
  - Maquina/modelo en aprendizaje por autonomía propia.
- **Acción**
  - Todas las acciones posibles que el '**Agente**' puede decidir efectuar/tomar.

- **Recompensas**
  - Premios otorgados al '**Agente**' al acertar y/o tomar el camino correcto.
- **Penalizaciones**
  - Castigos impuestos al '**Agente**' por fallar y/o tomar el camino incorrecto.

Upper Confidence Bound (UCB)

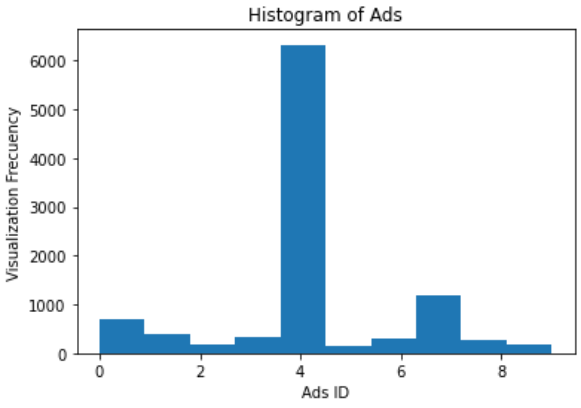
Algoritmo que va ajustando un '**rango de confianza**', que va buscando/ encerrando la '**Distribución Media**' de cada variable del DataSet, tratando de hacer converger ambos estadísticos (R.C. y D.M.). Así, logra determinar la variable más significativa.

Costoso por requerir ajustes en los estadísticos en cada iteración.

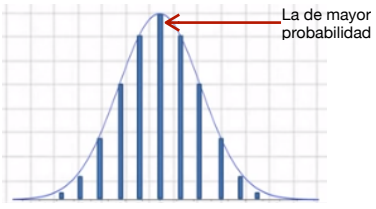


Algoritmo determinista (a mismas entradas, misma secuencia de estados internos, por ende, siempre la misma salida).

A cada '**iteración**' se re-definen los '**estadísticos**' (D.M. y R.C.).



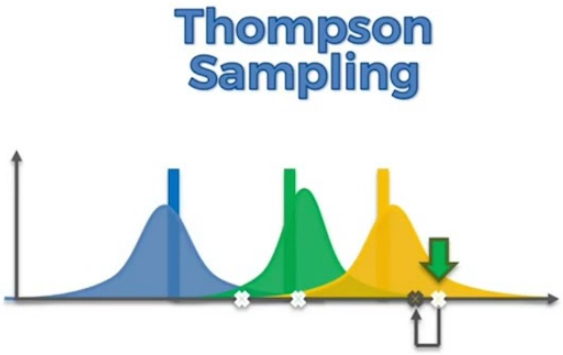
Distribución de la Media
*Conjunto de '**medias**' (promedios) que se pueden encontrar en los datos y la probabilidad de cada una de ser seleccionada.*



Muestreo Thompson

Algoritmo que evalúa los datos de cada variable del DataSet de manera aleatoria hasta que identifica la de mayor '**Distribución Media**', entonces, continua solo evaluando esa variable hasta dar con su mejor media.

Resultados superiores a UCB, flexible y permite acumular datos para una posterior explotación.



Algoritmo estocástico (a mismas entradas, secuencia de estados internos aleatorio, por ende, salida diferente/aleatoria).

A diferencia de **UCB**, este algoritmo busca '**acercarse**' a la media real de la variable; no identificarla.