A multi armed bandit is known in literature as a problem of a trade-off between gaining information from the system and exploiting the already available information. The name comes from playing a slot machine, where one pulls the handle (arm) and wins with certain fixed probability, not known to the player beforehand. The multi armed case consists in a set of such slot machines, each having their own fixed but unknown probabilities of a win. By pulling the machines' handles a player wins or looses and gains information on how prone each machine is to success. As in any game of chance, the goal is to maximise the number of wins as one keeps pulling the machines' handles.

Since in the game one wants to maximise profits, the previous experience with a slot machine continuously updates players expectations of how probable it is to win at it. This falls into the concept of Bayesian interpretation of probability: a likelihood of success is a measure of belief, where the belief is based on the previous experience. Intuitively, the more we experiment, the better we estimate our chances.

In statistics an experiment with two possible outcomes (1 for success, 0 - failure) is commonly called a Bernoulli trial. The single trial obeys a so-called Bernoulli distribution - a discrete probability distribution of a variable, which takes a value of success with a certain probability or failure otherwise. In case the probability is not fixed but rather randomly drawn from a bell-shaped probability density function, skewed towards the true probability of success, such trial obeys the Beta-Bernoulli distribution. Here Beta stands for Euler's Beta-function, which actually is shaped like a skewed bell.

An effective way to face the multi armed bandit problem is to use a Thompson model. The player chooses the slot machine based on his expectations of success at each of the machines. The expectation of success is modelled with the Beta-Bernoulli distribution, exploiting previous experience. So, in a sense, the bell-shaped Beta-Bernoulli distribution reflects the uncertainty of a player of the exact probability to win. The algorithm randomly samples from Beta distributions of each slot machines, and chooses the machine with the maximal sampled value in each round. New player's action updates his expectation based on the outcome of the action. The more evidence there is, that a particular machine has higher success rate, the more likely this machine will be chosen for the next action. However due to using Beta-distributions, other machines keep being tried once in a while due to statistical chance and depending on their performance in the past.