

Input: Set of slot machines S , exploration parameter ϵ

Initialization: $R(s) \leftarrow 0$ for all $s \in S$;

Initialization: $N(s) \leftarrow 0$ for all $s \in S$;

for $t = 1$ *to* T **do**

if $random(0, 1) > \epsilon$ **then**

 Choose action $a_t = \arg \max_{s \in S} R(s)$;

else

 Choose a random action $a_t \sim Uniform(S)$;

end

 Perform action a_t on the chosen slot machine;

 Observe reward r_t obtained from the chosen action;

 Update action count: $N(a_t) \leftarrow N(a_t) + 1$;

 Update estimated reward value: $R(a_t) \leftarrow \frac{R(a_t) \cdot (N(a_t) - 1) + r_t}{N(a_t)}$;

end