Mathematical Problem – Direct Utility Estimation (Tourist Example)

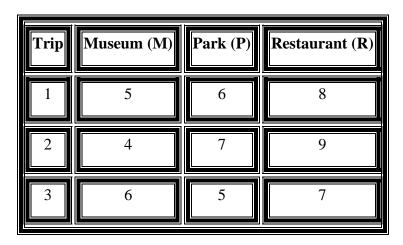
A tourist is visiting three places in a city:

Museum (M), Park (P), Restaurant (R).

The tourist follows a **fixed policy**:

- Always visits places in this order: Museum \rightarrow Park \rightarrow Restaurant.
- At each place, the tourist receives an **enjoyment score** (**reward**).
- The goal is to estimate the **Direct Utility** (**U**) **of each place** based on multiple visits.

Given Data (Rewards for 3 trips)



The utility U(s)U(s) of a place is the **average reward** received when visiting that place.

Step-by-Step Solution (Direct Utility Estimation)

We calculate the **average reward** for each place:

Step 1: Calculate the Direct Utility for Each Place

$$U(M) = rac{ ext{Total Reward at Museum}}{ ext{Number of Visits}} \ U(P) = rac{ ext{Total Reward at Park}}{ ext{Number of Visits}} \ U(R) = rac{ ext{Total Reward at Restaurant}}{ ext{Number of Visits}}$$

Step 2: Compute the Values

$$U(M) = rac{5+4+6}{3} = rac{15}{3} = 5.0$$
 $U(P) = rac{6+7+5}{3} = rac{18}{3} = 6.0$ $U(R) = rac{8+9+7}{3} = rac{24}{3} = 8.0$

Final Answer: Estimated Utilities

- Museum: U(M)=5.0
- ullet Park: U(P)=6.0
- Restaurant: U(R)=8.0

These values represent the **estimated enjoyment of each place**, based purely on past experiences. However, this method **does not consider** how places connect or affect future experiences (which ADP and TD Learning would do).