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 → Park → 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)

Trip	Museum (M)	Park (P)	Restaurant (R)
1	5	6	8
2	4	7	9
3	6	5	7

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

$$U(M) = \frac{\text{Total Reward at Museum}}{\text{Number of Visits}}$$

$$U(P) = \frac{\text{Total Reward at Park}}{\text{Number of Visits}}$$

$$U(R) = \frac{\text{Total Reward at Restaurant}}{\text{Number of Visits}}$$

Step 2: Compute the Values

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

Final Answer: Estimated Utilities

- Museum: $U(M) = 5.0$
- 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). 🚀