# **Mathematical Problem – Direct Utility Estimation (Tourist Example)**

# **Solving the Tourist Problem Using Adaptive Dynamic Programming (ADP)**

In Adaptive Dynamic Programming (ADP), we:

- ✓ Build a model of the environment (state transitions and rewards).
- ✓ Use the **Bellman equation** to compute the utilities of each place.
- ✓ Continuously refine utility estimates using the environment model.

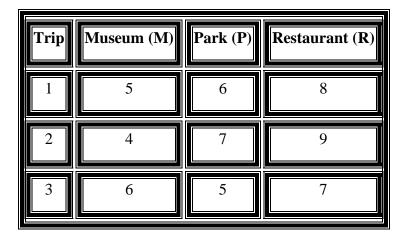
# **Problem Setup:** A **tourist** visits three places in a city:

**Museum (M),** ♠ Park (P), ♠ Restaurant (R)

### **Fixed policy**:

- The tourist always visits places in this order:  $Museum \rightarrow Park \rightarrow Restaurant$
- Each place provides a **reward** (enjoyment score).
- The tourist wants to estimate the utility of each location based on both immediate rewards and future rewards.

### **Given Rewards (Per Trip)**



We will use the **Bellman equation** to find the utility of each place.

#### **Step 1: Define the Bellman Equation**

The **utility** U(s) of a place depends on:

- ✓ **Immediate reward r(s)** (enjoyment at that place).
- Future rewards from the next place.

The Bellman equation is:

$$U(s) = r(s) + \gamma U(s')$$

where:

- U(s) = Utility of current place.
- r(s) = Immediate reward at the current place.
- $\gamma$ \gamma = Discount factor (importance of future rewards, typically 0.90.9).
- U(s') = Utility of the next place.

#### Step 2: Initialize Rewards and Transition Probabilities

We estimate the average rewards (same as in Direct Utility Estimation):

$$r(M) = 5.0, \quad r(P) = 6.0, \quad r(R) = 8.0$$

Since the tourist always follows the same path, the transitions are:

- Museum  $\rightarrow$  Park  $\rightarrow$  Restaurant
- The final state (**Restaurant**) has **no future place**, so U(R) = r(R).

## **Step 3: Compute Utilities Using the Bellman Equation**

1. **Utility of the Restaurant U(R)** (Final Place):

$$U(R) = r(R) = 8.0$$

2. Utility of the Park U(P):

$$U(P)=r(P)+\gamma U(R)$$
  $U(P)=6.0+(0.9 imes 8.0)$ 

U(P) = 6.0 + 7.2 = 13.2

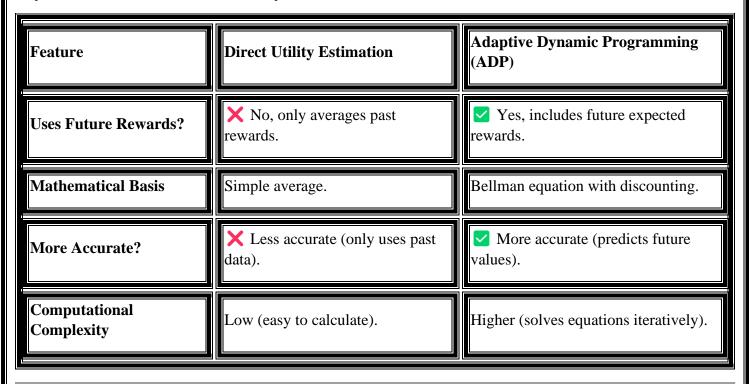
3. Utility of the Museum U(M):

$$U(M) = r(M) + \gamma U(P)$$
  $U(M) = 5.0 + (0.9 imes 13.2)$   $U(M) = 5.0 + 11.88 = 16.88$ 

### **Final Answer: Estimated Utilities Using ADP**

Place	<b>Direct Utility Estimation U(s)</b>	ADP U(s) (Bellman Equation)
Museum (M)	5.0	16.88
Park (P)	6.0	13.2
Restaurant (R)	8.0	8.0

# **Key Differences: ADP vs. Direct Utility Estimation**



#### Conclusion

✓ ADP gives better estimates because it considers future rewards instead of just past data.

# **Real-life Example:**

- **Direct Utility Estimation:** The tourist **only remembers past experiences** and rates places based on past visits.
- **ADP:** The tourist **predicts** future experiences based on how places are connected (e.g., a park near a great restaurant is **more valuable**).

ADP is more powerful because it helps make smarter travel decisions by considering the long-term value of each location.

**Decision: Which Place is Better?** 

- Best Place to Start = Museum (M) → Utility = 16.88
  - The Museum is the best place to start **because it leads to high-value future rewards** (Park → Restaurant).
  - It means that the Museum not only has good rewards but also leads to better places later.
- Second Best Place = Park (P)  $\rightarrow$  Utility = 13.2
  - The Park is valuable but **not as much as the Museum**, because it only leads to the Restaurant.
- Least Valuable Place = Restaurant (R)  $\rightarrow$  Utility = 8.0
  - The Restaurant has no future rewards since it is the last stop.

**Final Conclusion: Where Should the Tourist Start?** 

**The tourist should start at the Museum (M)** because it has the **highest utility (16.88)**, meaning it provides the most **long-term enjoyment**.



- It leads to the Park, which has a good future value.
- The Park then leads to the Restaurant, which gives the final reward.
- This sequence maximizes overall satisfaction!