Question 1

Yes, there is a difference in the context in which transition probabilities are used in a Markov Decision Process (MDP) and a Partially Observable Markov Decision Process (POMDP), although the fundamental concept of transition probabilities remains the same.

In an MDP, the transition probabilities define the likelihood of transitioning from one state to another given a particular action. This model assumes full observability of states, meaning the current state is always known when making a decision.

In contrast, a POMDP extends the MDP framework to handle situations where the states are not fully observable. Instead of operating on states directly, a POMDP works with beliefs or probability distributions over states. Transition probabilities in a POMDP still describe the dynamics of the underlying process, but decisions are made based on the belief state rather than on the actual state, which is not directly observable.

So, while the nature of the transition probabilities is fundamentally the same (they both express how the system evolves), their application differs due to the additional layer of complexity introduced by partial observability in POMDPs.

Therefore, the correct answer to the question is:
True

Question 2

If the observation function in a POMDP always gives you the exact state, then the POMDP effectively becomes an MDP because the uncertainty about the state is removed. In a standard POMDP, the agent receives observations that may provide only partial information about the true state due to the system's inherent uncertainty. However, if the observation function is perfect and always provides the exact state, the "partially observable" aspect of the POMDP is negated.

So, with perfect (exact) state observations, the agent in a POMDP will always know the true state, making the decision process identical to that in an MDP, where the state is assumed to be fully observable at all times. Under these conditions, the belief state is no longer a probability distribution over possible states but collapses to a certainty about the current state, which aligns with the definition of an MDP.

Therefore, the correct answer to the question is:
True

Question 3

In the case of POMDPs, the policy is not a direct mapping from states to actions because the states are not fully observable. Instead, a policy in a POMDP maps belief states to actions. A belief state represents a probability distribution over all possible states given the history of actions and observations. This is a key distinction between MDPs and POMDPs.

So the correct answer to the question is:
False

In a POMDP, the policy is a mapping from belief states (probability distributions over states) to actions, not from actual states to actions as in an MDP.

Question 4

In the context of a POMDP, a sufficient statistic is a minimal set of information that is needed to make an optimal decision. In a POMDP, since the states are not fully observable, the agent must rely on a history of actions and observations to make decisions.

a. History of actions and observations: This is indeed a sufficient statistic for a POMDP because the history of actions and observations can be used to infer the belief state, which is a probability distribution over states.

c. Probability distribution over states: This refers to the belief state, which is the primary sufficient statistic in a POMDP. The belief state encapsulates all relevant information from the history of actions and observations and is used to make optimal decisions.

b. History of actions: On its own, the history of actions is not sufficient because, without the corresponding observations, the agent cannot infer the belief state.

d. Previous state: Knowledge of the previous state alone is not sufficient for optimal decision-making in a POMDP because it does not provide information about the current state's probability distribution without observations.

Therefore, the correct answers are:
a. History of actions and observations
c. Probability distribution over states

Question 5

a. Policies can gather information and thus increase chances of getting reward later: This is a potential difference when using a POMDP model. In POMDPs, policies can be designed not just to take immediate rewarding actions, but also to gather information that may improve the agent's understanding of the state (belief state), which can lead to better decision-making and potentially higher rewards in the future. This aspect is unique to POMDPs due to their partially observable nature and is not typically a consideration in MDPs, where the state is fully observable.

b. The policy does not yield negative rewards: This statement is not necessarily a difference between POMDPs and MDPs. Both POMDPs and MDPs can have policies that yield negative rewards because the rewards are a function of the state and action, and do not inherently depend on whether the state is fully observable (MDP) or partially observable (POMDP). Negative rewards are used to model costs or penalties and can be present in both types of decision-making models.

The correct answer is:
a. Policies can gather information and thus increase chances of getting reward later

Question 6

The dimensionality of a belief in a POMDP depends on the number of states in the state space, not on the number of actions or observations. A belief state is a probability distribution over all possible states. Therefore, for each state, there is a corresponding belief about the probability of being in that state. The size of this belief vector is directly proportional to the number of states.

a. Number of actions: This does not affect the dimensionality of a belief. Actions may influence the update of the belief state, but they do not determine its size.

b. Number of observations: Like actions, observations influence how the belief state is updated (since new observations are used to refine the belief), but they do not determine the size of the belief state itself.

c. Number of states: This is what determines the dimensionality of the belief state. If there are N states, the belief state will be an N-dimensional vector where each dimension corresponds to the probability of being in one of the N states.

The correct answer is:
c. Number of states

Question 7

The QMDP (Quick Markov Decision Process) algorithm is an approximate solution method for POMDPs. It simplifies the POMDP by assuming that after the current decision, the system will become fully observable. This means that the QMDP value function is computed by considering the expected total reward starting from the current belief state and acting optimally as if the system were to become fully observable after the current action.

This simplification generally leads to an optimistic estimate of the value function because it assumes that uncertainty about the state will be resolved in the future, which is not always the case in a POMDP. Therefore, the QMDP value function does not necessarily provide a lower bound for the POMDP value function; it may overestimate it.

Hence, the correct answer to the question is:
False

The POMDP value function reflects the best possible performance under uncertainty, while the QMDP value function reflects the performance under the assumption that the uncertainty is resolved in the next step, which can lead to higher expected rewards than what might actually be achievable under continuous uncertainty.

Question 8

To compute the next belief in a POMDP, you need:

b. Current belief, action: You start with the current belief (a probability distribution over states) and update it based on the action taken.

c. Current belief, action, observation: After taking an action, you get an observation. The belief state is updated by applying the observation model, which uses the latest observation, along with the current belief and the action taken, to produce the next belief state.

You do not need:

a. Reward, action, observation: The reward is not necessary for updating the belief state. It is used to evaluate the quality of actions taken from a particular state or belief state, but it does not influence the belief update directly.

The correct answers are:
b. Current belief, action
c. Current belief, action, observation

Question 9

True

Monte Carlo Tree Search (MCTS) can indeed be used in POMDP problems. MCTS is a heuristic search algorithm for some decision processes like POMDPs. It builds a search tree by using random samples of the search space and can be adapted to handle the uncertainty and partial observability in POMDPs through the use of belief states rather than true states.

Question 10

False

In a POMDP, compared to an MDP, it is generally more difficult to create a full search tree to find the best action. This is due to the additional complexity introduced by partial observability, which requires maintaining and updating a belief state—a probability distribution over all possible states. This added complexity makes the search space much larger and more computationally intensive to navigate compared to the search space of an MDP, where the state is fully observable.