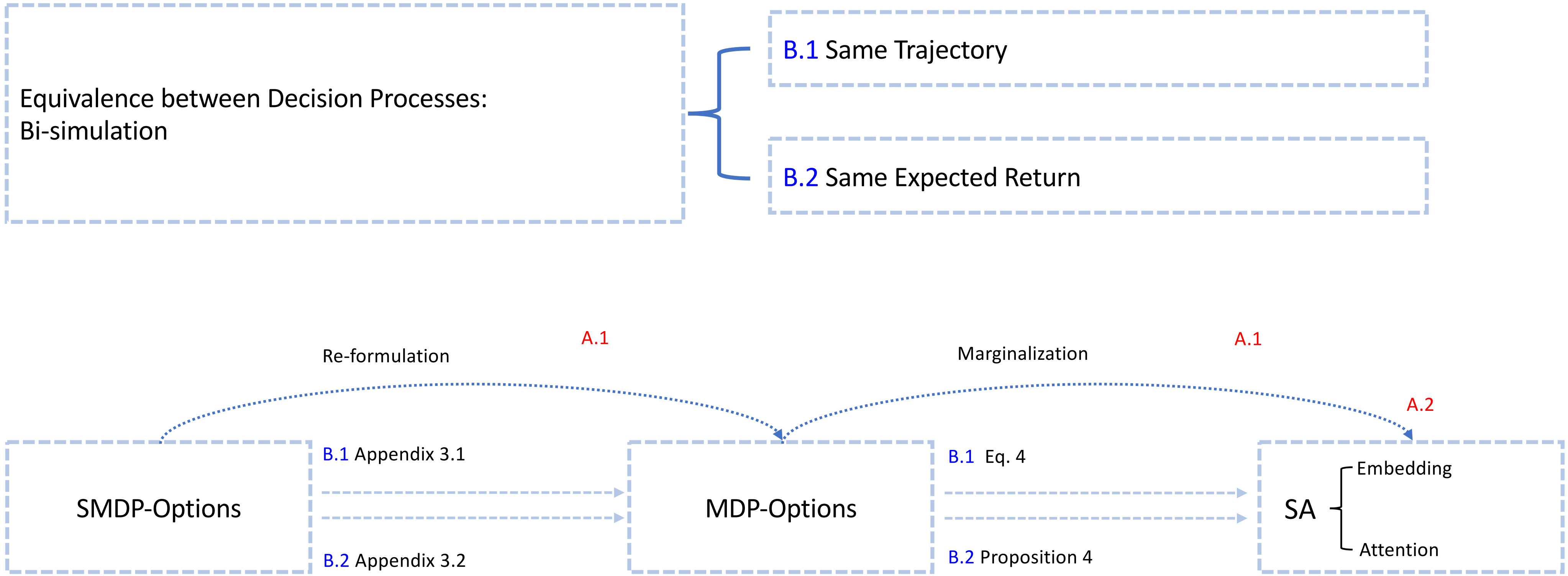
1. Equivalence: SMDP Option -> MDP Option -> SA

2. Option2Vec Embedding: HMM-like MDP -> Option Embeddings

e.g. **Learning** mixture distributions -> **Inference** latent variables

3. Transformer Decoder Implementation

1. Equivalence: SMDP Option -> MDP Option -> SA



SMDP Option

$$P(\tau) = P(\mathbf{s}_0)P(\mathbf{o}_0)P_{o_0}(\mathbf{a}_0|\mathbf{s}_0)\prod_{t=1}^{\infty} P(\mathbf{s}_t|\mathbf{s}_{t-1},\mathbf{a}_{t-1})P_{o_t}(\mathbf{a}_t|\mathbf{s}_t)$$

$$[P_{o_{t-1}}(\mathbf{b}_t = 0|\mathbf{s}_t)\mathbf{1}_{\mathbf{o}_t=o_{t-1}} + P_{o_{t-1}}(\mathbf{b}_t = 1|\mathbf{s}_t)P(\mathbf{o}_t|\mathbf{s}_t)].$$

Reformulate to HMM-like MDP

通过Hidden variable $\bar{O}$,
我们表示成了一个Mixture Distribution

此处每个下角标 P_o， 都是一个Distribution (神经网络)

$$\boxed{P(\mathbf{a}_t|\mathbf{s}_t,\bar{\mathbf{o}}_t)} = \prod_{o\in\bar{\mathbf{o}}_t}\boxed{P_o(\mathbf{a}_t|\mathbf{s}_t)^o,} \qquad P(\mathbf{b}_t|\mathbf{s}_t,\bar{\mathbf{o}}_{t-1}) = \prod_{o\in\bar{\mathbf{o}}_{t-1}} P_o(\mathbf{b}_t|\mathbf{s}_t)^o \tag{4}$$

$$P(\bar{\mathbf{o}}_t|\mathbf{s}_t,\mathbf{b}_t,\bar{\mathbf{o}}_{t-1}) = P(\bar{\mathbf{o}}_t|\mathbf{s}_t)^{\mathbf{b}_t}P(\bar{\mathbf{o}}_t|\bar{\mathbf{o}}_{t-1})^{1-\mathbf{b}_t}, \tag{5}$$

$$P(\tau/\bar{B}) = P(\bar{\tau}/\bar{B}) = P(\mathbf{s}_0)P(\bar{\mathbf{o}}_0)P(\mathbf{a}_0|\mathbf{s}_0,\bar{\mathbf{o}}_0)\prod_{t=1}^{\infty} P(\mathbf{s}_t|\mathbf{s}_{t-1},\mathbf{a}_{t-1})P(\mathbf{a}_t|\mathbf{s}_t,\bar{\mathbf{o}}_t)$$

MDP Option

$$\sum_{\mathbf{b}_t} P(\mathbf{b}_t|\mathbf{s}_t,\bar{\mathbf{o}}_{t-1})P(\bar{\mathbf{o}}_t|\mathbf{b}_t,\mathbf{s}_t,\bar{\mathbf{o}}_{t-1}) \tag{6}$$

MDP Option

$$P(\tau/\bar{B}) = P(\bar{\tau}/\bar{B}) = P(\mathbf{s}_0)P(\bar{\mathbf{o}}_0)P(\mathbf{a}_0|\mathbf{s}_0,\bar{\mathbf{o}}_0)\prod_{t=1}^{\infty}P(\mathbf{s}_t|\mathbf{s}_{t-1},\mathbf{a}_{t-1})P(\mathbf{a}_t|\mathbf{s}_t,\bar{\mathbf{o}}_t)$$
$$\boxed{\sum_{\mathbf{b}_t}P(\mathbf{b}_t|\mathbf{s}_t,\bar{\mathbf{o}}_{t-1})P(\bar{\mathbf{o}}_t|\mathbf{b}_t,\mathbf{s}_t,\bar{\mathbf{o}}_{t-1})} \tag{6}$$

Termination Variable b_t
被Marginalize掉了

Marginalization

SA

$$P(\bar{\tau}) = P(\mathbf{s}_0)P(\bar{\mathbf{o}}_0)P(\mathbf{a}_0|\mathbf{s}_0,\bar{\mathbf{o}}_0)\prod_{t=1}^{\infty}P(\mathbf{s}_t|\mathbf{s}_{t-1},\mathbf{a}_{t-1})P(\mathbf{a}_t|\mathbf{s}_t,\bar{\mathbf{o}}_t)P(\bar{\mathbf{o}}_t|\mathbf{s}_t,\bar{\mathbf{o}}_{t-1}) \tag{7}$$

2. Option2Vec Embedding

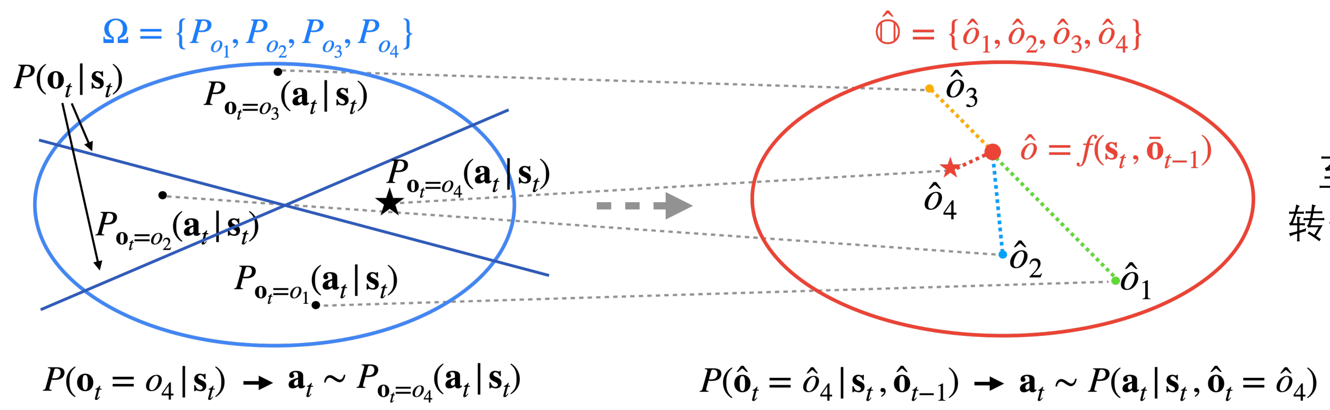
$$P(\bar{\tau}) = P(\mathbf{s}_0)P(\bar{\mathbf{o}}_0)P(\mathbf{a}_0|\mathbf{s}_0, \bar{\mathbf{o}}_0)\prod_{t=1}^{\infty} P(\mathbf{s}_t|\mathbf{s}_{t-1}, \mathbf{a}_{t-1})P(\mathbf{a}_t|\mathbf{s}_t, \bar{\mathbf{o}}_t)P(\bar{\mathbf{o}}_t|\mathbf{s}_t, \bar{\mathbf{o}}_{t-1}) \tag{7}$$

$$\boxed{P(\mathbf{a}_t|\mathbf{s}_t, \bar{\mathbf{o}}_t) = \prod_{o \in \bar{\mathbf{o}}_t} P_o(\mathbf{a}_t|\mathbf{s}_t)^o}$$

这里原本是Mixture Disbribution，
对于4个option仍然有4个distribution需要学习

$$P(\mathbf{a}_t|\mathbf{s}_t, \bar{\mathbf{o}}_t)$$

然而我们完全可以将O作为Hidden Variable，
扩展为Embedding Vector
因此对于4个Option，我们只有4个
Embedding Vector 需要学习



至此，我们将Learning 4个 P_o 的问题，
转化为Learning 4个 Embedding Vector $\bar{O}$ ，
并从中Inference $\bar{O}$ 的问题

3. Transformer Decoder Implementation