Active Inference in continuous time notes

Federico Maggiore

September 25, 2020

Summary notation

- $\mathbf{x} = \{x_i\}_{i=1}^{D}$ environmental variables of the D-dimensional space constituting latent or hidden states;
- $\mathbf{s} = \{s_i\}_{i=1}^{S}$ body sensors input;
- $P(\mathbf{s}, \mathbf{x})$ *G-density*;
- $P(\mathbf{x}|\mathbf{s})$ *Posterior*;
- $P(\mathbf{s}|\mathbf{x})$ *Likelihood*;
- $P(\mathbf{x})$ *Prior*;
- $P(\mathbf{s}) = \int P(\mathbf{s}|\mathbf{x})P(\mathbf{x})d\mathbf{x}$ *marginal likelihood*
- $Q(\mathbf{x})$ *R-density*
- $F \equiv \int Q(\mathbf{x}) \ln \frac{Q(\mathbf{x})}{P(\mathbf{x},\mathbf{s})} d\mathbf{x}$ *Variational Free Energy*

Free Energy Principle (FEP)

The goal of an agent is to determine the probability of the hidden states given some sensory inputs:

$$P(\mathbf{x}|\mathbf{s}) = \frac{P(\mathbf{s}, \mathbf{x})}{P(\mathbf{s})} = \frac{P(\mathbf{s}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{s})} \tag{1}$$

with

- $P(\mathbf{s}, \mathbf{x})$ *G-density*, beliefs about the states assumed to be encoded by the agent;
- $P(\mathbf{x}|\mathbf{s})$ *Posterior*, i.e. probability of hidden causes x given observed sensory data;
- $P(\mathbf{s}|\mathbf{x})$ *Likelihood*, i.e. organism's assumptions about sensory input $\mathbf{s}$ given the hidden causes $\mathbf{x}$;
- $P(\mathbf{x})$ *Prior*, i.e. agent's beliefs about hidden causes **before** that $\mathbf{s}$ are received;

- $P(\mathbf{s}) = \int P(\mathbf{s}|\mathbf{x})P(\mathbf{x})d\mathbf{x}$ *marginal likelihood*, i.e. normalization factor.

For the agent it's not necessary to compute the complete posterior distribution, it has only to find the hidden state -or at least a good approximation- that maximize the posterior, i.e. $\arg\max_{\mathbf{x}} P(\mathbf{x}|\mathbf{s})$. The problem with the exact Bayesian scheme, is that $P(\mathbf{s})$ is often impossible to calculate, and moreover $P(\mathbf{x}|\mathbf{s})$ may not take a standard shape and could not have a summary statistics.

A biologically plausible technique consist in using an auxiliary distribution $Q(\mathbf{x})$ called *recognition density* (*R-density*) that has to be optimized to became a good approximation of the posterior.

In order to do this the Kullback-Leibler divergence is minimized:

$$\begin{aligned} D_{KL}(\, Q(\mathbf{x}) \,||\, P(\mathbf{x}|\mathbf{s})\,) &= \int Q(\mathbf{x}) \ln \frac{Q(\mathbf{x})}{P(\mathbf{x}|\mathbf{s})} d\mathbf{x} \\ &= \int Q(\mathbf{x}) \ln \frac{Q(\mathbf{x})P(\mathbf{s})}{P(\mathbf{x},\mathbf{s})} d\mathbf{x} \\ &= \int Q(\mathbf{x}) \ln \frac{Q(\mathbf{x})}{P(\mathbf{x},\mathbf{s})} d\mathbf{x} + \ln P(\mathbf{s}) \int Q(\mathbf{x}) d\mathbf{x} \\ &= F + \ln P(\mathbf{s}) \end{aligned} \tag{2}$$

where

- $F \equiv \int Q(\mathbf{x}) \ln \frac{Q(\mathbf{x})}{P(\mathbf{x},\mathbf{s})} d\mathbf{x} = -\int Q(\mathbf{x}) \ln P(\mathbf{x},\mathbf{s}) d\mathbf{x} + \int Q(\mathbf{x}) \ln Q(\mathbf{x}) d\mathbf{x}$ is the *Variational Free Energy*, a quantity that depends on the R-density and the knowledge about the environment i.e. the G-density $P(\mathbf{s},\mathbf{x}) = P(\mathbf{s}|\mathbf{x})P(\mathbf{x})$ that we are assuming the agent has.
- $\ln P(\mathbf{s})$ is a term independent of the recognition density $Q(\mathbf{x})$ ($\Rightarrow$ minimizing F with respect to $Q(\mathbf{x})$ will minimize the D_{KL})

Laplace approximation

Often optimizing F for arbitrary $Q(x)$ is particularly complex. Moreover, assumed that neural activity parametrise sufficient statistic, a common approximation is to assume that

Given a probability density $Q(x)$ with a peak at point μ, the Taylor-expansion of the logarithm around his peak is

$$\ln Q(x) \simeq \ln Q(\mu) - \frac{1}{2} \frac{(x-\mu)^2}{\Sigma} \tag{3}$$

with

$$\frac{1}{\Sigma} = -\frac{\partial^2}{\partial x^2} \ln P(x) \bigg|_{x=\mu} \tag{4}$$

Now it is possible to approximate the probability distribution $Q(x)$ with the distribution

$$\mathcal{N}(x;\mu,\Sigma) = \frac{1}{\sqrt{2\pi\Sigma}}\, e^{\frac{(x-\mu)^2}{2\Sigma}} \tag{5}$$

a Gaussian distribution that has been normalized using the factor

$$Z_{\mathcal{N}} = Q(\mu)\sqrt{2\pi\Sigma} \tag{6}$$

Generalizing for a density $Q(\mathbf{x})$ over a K-dimensional space $\mathbf{x}$ with peak at $\boldsymbol{\mu}$

$$\ln Q(\mathbf{x}) \simeq \ln Q(\boldsymbol{\mu}) - \frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \tag{7}$$

with

$$\left[\boldsymbol{\Sigma}^{-1}\right]_{i,j} = - \frac{\partial^2}{\partial x_i \partial x_j} \ln P(\boldsymbol{x}) \bigg|_{\mathbf{x}=\boldsymbol{\mu}} \tag{8}$$

is possible to approximate $Q(\mathbf{x})$ with the multivariate Gaussian distribution

$$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^K \det \boldsymbol{\Sigma}}} e^{\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})} \tag{9}$$

This approximation is particularly useful to approximate integrals.