能動的推論攻略ドキュメント v0.3

アジェンダ：
- **期待自由エネルギーの導出が書きかけ**
- **プロセスモデルの具体例がほしい**
- **Appendixが現状雑魚すぎるのでわかりやすくする**
- **イントロ的な部分の強化**

このドキュメントは、Fristonの能動的推論を0から理解することを目的に作られた。
説明のスタイルはTschantz et al. (2020) 並びにSmith et al. (2021) を大いに参考にしている。

自由エネルギー原理は、生きているシステムが満たさなければいけない条件を主張する規範的な主張である。現実世界に存在するものは、無秩序に向かっていく自然な傾向を持つ。生物とそれ以外の物理的存在物を分けるのは、その自然な傾向に抗うシステムであるかどうかである。無秩序に向かうのに抗することは、数学的にはエントロピーを下げることに相当する。それは自由エネルギーを最大化することで実現できる。従って、自由エネルギー原理とは：

物理的存在物が有している、無秩序に向かっていく自然な傾向へ抗うために、生物は自由エネルギーの最小化を通じてエントロピーを環境へ排出している

という主張である (Friston, 2010)。まだ、「エントロピー」も「自由エネルギー」も定義していないので、なんのこっちゃという感じだが、ひとまずそういう主張であるということは念頭に置いておこう。自由エネルギー原理を用いて何か具体的な現象を説明・予測をするには、原理から何か「プロセス理論 (process theory)」を導出しなければいけない (Friston et al., 2017)。プロセス理論に相当するのが、個別の確率モデルになってくる。それら個別の理論が満たすべき条件を提唱しているため「原理」という言葉を使っているわけだ。

当初、自由エネルギー原理は、知覚を対象とした皮質の計算原理として提唱された (Friston, 2005)。その後、行動や学習、感情、運動制御と広く適用できることを示していった。
最近のFristonはさらに、生命の満たすべき条件とまでスケールを広げていっている (Ramstead et al., 2018)。とにかくまあ言ってることがでかいので、賛成するにせよ反対するにせよ、知っておいて損はない。

問題は、Fristonオリジナルの論文は数学的にも、英文的にも、読むのに困難を極める点にある。最近になって、Tschantz et al. (2020) や、Smith et al. (2021)、Bogacz (2017) あるいは日本神経回路学会の自由エネルギー原理特集号 (島崎他, 2018) のような解説が出てきて、アクセスしやすくはなった。しかしそれでもなお、心理学者の平均的な数学能力を超えているものが多い。例外的に吉田・田口 (2018) は平易に自由エネルギー原理のエッセンスを紹介しているが、ある程度単純化がなされているため、そこから他の論文に移る際にギャップが大きい。そこで、このドキュメントは、基本的に事前知識がなく読めることを目標としつつ、本格的な能動的推論

の解説を行う。仮定する数学的知識は、「対数の積は和になる ($\log ab = \log a + \log b$)」くらいに留める。その他必要な知識は、その都度説明するか、Appendixにつけた。

1. 問題設定

我々は動物は感覚器官を通じて、何らかの観測 o を得ている。観測 o は、問題設定次第でなんでもよい。例えば、食物の大きさかもしれないし、光の強さかもしれない、報酬の有無でもよい。その観測は、隠れ状態 x から生み出されている。隠れ状態 x も問題設定により決まる。例えば o が網膜像の大きさなら、x はそこに映る物体の真の大きさと考えるのがよいだろう。 o が気になる異性からの連絡であれば、x はその異性のあなたへの気持ちとしてモデル化するのがよいだろう。そのように何が o で何が x かは、扱っている状況よりけりということだ。

自由エネルギー原理では、エージェントは、観測 o と隠れ状態 x についてのモデルを有している。そのモデルを生成モデル (generative model) という。生成モデルは、o と x の同時分布 $p(x, o)$ の形で表現される。生成モデルはあとで役に立つから、ここで概念だけは導入しておく。

いきなり天下りに生成モデルが出てきて「生物は観測 o と隠れ状態 x の同時分布を生成モデルとして有している」なんて言われも・・・と思うかもしれない。そこで、こう考えたらわかりやすい。生物は、感覚入力の原因 x の下で、実際の入力に相当する o がどれくらい起きやすいかを評価できる。つまり、尤度分布 $p(o\,|\,x)$ を、意識・無意識的な形で評価している。「昨日雨が降った (x) 下で、窓が濡れているのがどれくらいありえそうか」をエージェントは考えられるという風に見たら、この主張は突飛ではないだろう。さらに、エージェントはその隠れ状態 x の確率分布 $p(x)$ も有している。この例の場合、「そもそも雨が降る確率はどれくらいか？」に相当するわけだが、その地域に長く住んでいれば、そこがどれくらい雨が降る地域なのか、概ねの感覚はあるだろう。$p(x)$ はこのように、生活を通じて個体発生的に獲得される場合もあるだろうし、進化を通じて生得的に獲得されている場合もあるだろう。ここで、$p(o\,|\,x)$ と $p(x)$ をエージェントが有しているのを認められれば、たちどころに次の等式が成り立つ。

$$p(o\,|\,x)p(x) = p(x, o)$$

なぜなら、条件付き確率の定義が $p(o\,|\,x) = \dfrac{p(x, o)}{p(x)}$ であるためだ。つまり、尤度分布 $p(o\,|\,x)$ と隠れ状態の分布 $p(x)$ をエージェントが有していると仮定したら、自動的にエージェントが生成モデル $p(x, o)$ を保持しているのを認めるのと同義なのだ。

さて、通常、本当に知りたいのは隠れ状態 x だ。例えば捕食者なら、網膜像の大きさ (o) から獲物の真の大きさ (x) を知りたいだろう。つまり、o から x を推論する必要がある。現実世界はノイジーなので、この推論は確率的なものになる。よって、$p(o\,|\,x)p(x) = p(x, o)$ というエージェントが

利用できる情報に基づいて、$p(x|o)$ を推論することがエージェントに課せられたタスクである。この条件付き確率は、事後分布としてベイズ則により次のように書ける。

$$p(x|o) = \frac{p(o|x)p(x)}{p(o)}$$

$p(o|x)$ は尤度、$p(x)$ は x の事前分布である。尤度は、ある隠れ状態 x の下で観測 o を得る確率を表す。例えば、気になる異性があなたに好意がある ($x = love$) という状態の下での連絡 $o = frequently$ の確率は高いだろう。言い換えれば、 $p(o = freqently|x = love)$ は高い値になるはずだ。一方、逆になんの脈もない ($x = do\ not\ like$) とき、連絡をたくさんもらうという観測が得られる確率は残念ながら低くなる ($p(o = freqently|x = do\ not\ like)$ はとても低い)。尤度とはそのような確率を表現している。事前分布 $p(x)$ は、各隠れ状態の事前確率を表現している。

$p(o)$ は周辺尤度、あるいはモデルエビデンスという名前がついている量である。$p(o)$ は、o と x の同時分布を周辺化したら得られる確率分布であるため

$$p(o) = \sum_x p(x, o) = \sum_x p(o|x)p(x)$$

である。つまり、すべての x について足し上げていかなければ、$p(o)$ を直接求めることはできない。x が連続量のときは、x の数だけ多重積分をすることになる。ベイズ統計学ではおきまりの導入なのだが、現実世界のデータでは、$p(o)$ はすぐに計算量が爆発してしまい、直接求めるのは不可能なことが多い。よって、事後分布 $p(x|o)$ は通常、解析的に求めることができない。

2. 自由エネルギーの導入

事後分布 $p(x|o)$ は直接評価ができないと言った。そこで、別の方針を考えるわけだが、役に立つのが変分推論の考え方だ。事後分布 $p(s|o)$ の代わりに近似事後分布 $q(x)$ を用意し、それを $p(x|o)$ に近づけていこうというのが変分推論の考え方である。近似事後分布 $q(x)$ は、自由エネルギー原理の文脈では、認識分布 (recognition density) と呼ばれる。事後分布 $p(x|o)$の近似に使うわけなので、$p(x|o)$ より単純で扱いやすい分布でなきゃ意味がない。どういう風に単純にするかは後に回すとして、ここでは $q(x)$ がパラメータ ϕ で完全に形が決まる分布であるということにしておこう[1]。なので本当は $q(x|\phi)$ とでも書くところなのだが、強調したいところ以外では省略しておく。ただ、$q(x)$ はパラメータ ϕ で形が決まる分布なのだということは念頭に置いておいてほしい。

---

[1] このようなパラメータを十分統計量 (sufficient statistics) という。例えば、正規分布の十分統計量は平均と分散である。

どうやって、$p(x|o)$ に $q(x)$ を近づけていくのか？そのためには両者がどれくらい異なるのかを測る尺度が必要になる。分布間の類似性を測るのに便利なのが、カルバック・ライブラー情報量(Kullback-Leibler divergence) である。カルバック・ライブラー情報量 D_{KL} は、二つの分布 $p(x)$ と $q(x)$ の間で、次のように定義される。

$$D_{KL}[q(x) \| p(x)] = \sum_x q(x) \log \frac{q(x)}{p(x)}$$

$D_{KL}[q(x) \| p(x)]$ という書き方はいかめしく見えるが、$p(x)$ と $q(x)$ の D_{KL} を測っているということを述べている。早速、カルバック・ライブラー情報量を使って事後分布 $p(x|o)$ と認識分布 $q(x)$ の間の類似性を表現してみよう。

$$D_{KL}[q(x) \| p(x|o)] = \sum_x q(x) \log \frac{q(x)}{p(x|o)}$$

せっかく書いたのはいいのだが、これは直接求めることはできない。なぜなら、式の中に $p(x|o)$ が入っているからだ。$p(x|o)$ が直接求められないのに、それが式に入っているのだから、当然このカルバック・ライブラー情報量も評価できない。このままではバカみたいなんだが、ここで諦めずに式を変形してみよう。$p(x|o) = \frac{p(x,o)}{p(o)}$ なので、

$$\begin{aligned}
D_{KL}[q(x) \| p(x|o)] &= \sum_x q(x) \log \frac{q(x)}{p(x|o)} \\
&= \sum_x q(x) \log \frac{q(x)p(o)}{p(x,o)} \\
&= \sum_x q(x) \left(\log \frac{q(x)}{p(x,o)} + \log p(o) \right) \\
&= \sum_x q(x) \log \frac{q(x)}{p(x,o)} + \sum_x q(x) \log p(o) \\
&= \sum_x q(x) \log \frac{q(x)}{p(x,o)} + \log p(o) \\
&= \mathbb{E}_{q(x)} \left[\log \frac{q(x)}{p(x,o)} \right] + \log p(o)
\end{aligned}$$

2段目から3段目の変形では、対数の積が和になる法則を使っている($\log ab = \log a + \log b$)。4段目から5段目では、確率分布の定義 $\sum_x q(x) = 1$ を使っている。$p(o)$ は x に依存しないため、和の外に出せてしまう。従って $q(x)$ が消えてしまっている。また、最終段は、単に期待値の定義に従って書き直しただけだ。ここで

$$F = \sum_x q(x)\log\frac{q(x)}{p(x,o)} = \mathbb{E}_{q(x)}\left[\log\frac{q(x)}{p(x,o)}\right]$$

と定義する。これが変分自由エネルギー F である。これで

$$D_{KL}[q(x)\|p(x|o)] = F + \log p(o)$$

となることがわかった。この変形には2つの美味しい点がある。第一に、思い出してほしいこととして、$D_{KL}[q(x)\|p(x|o)]$ は、$q(x)$ を変化させることによって変えられるということだ。つまり、$q(x)$ のパラメータ ϕ を実際には最適化していくわけだが、$\log p(o)$ はそれとは無関係な量である。つまり、$D_{KL}[q(x)\|p(x|o)]$ の最小化に関係するのは F である。F を下げれば、$D_{KL}[q(x)\|p(x|o)]$ も下がる。よって、F を最小化すれば $D_{KL}[q(x)\|p(x|o)]$ の最小化という目的が達成されるわけだ。$p(x,o)$ はエージェントが持つ生成モデルであるため、形がわかっている。認識分布 $q(x)$ は、いわばエージェントが持つ現在の x についての推測である。$q(x)$ も $p(x,o)$ も形がわかっているので、F は計算できる。

さらに、カルバック・ライブラー情報量が非負であることを用いれば

$$D_{KL}[q(x)\|p(x|o)] = F + \log P(o) \geq 0$$
$$F \geq -\log P(o)$$

という不等式が現れてくる。これは何を意味しているのか？$-\log p(o)$ は、観測 o の情報量になっている。$-\log p(o)$ が高いほど、その観測 o は得られにくいものであったということである。そこで、$-\log p(o)$ は「サプライズ (surpise、あるいはsurprisal)」と呼ばれている。また、ベイズ統計学では、$-\log p(o)$ は (逆温度 $\beta = 1$ の) ベイズ自由エネルギーと呼ばれる量でもある。上の不等式は、サプライズが F より大きくはならないということを意味している。このことを、「F はサプライズの上限 (upper bound) を決めている」という。また、論文によっては逆に言い方をすることもある。その場合は $-F < \log p(o)$ を使って、「負の自由エネルギーはエビデンスの下限 (lower bound) となっている」と表現できる。いずれにせよ、言っていることは同じで、F を下げると、観測のサプライズの取りうる値の上限が下がっていくわけだ。つまり、観測のサプライズは平均的に下がっていく。

$p(o) = \sum_x p(x,o) = \sum_x p(o|x)p(x)$であるため、ありうるすべての x について足し上げた観測 o の確率分布である。観測 o が得られる確率なので、どんな感覚情報が得られるかの「期待」を $p(o)$ は表現している。対数は単調増加関数なので、その負の値を取ると意味が逆転する。つまり、$-\log p(o)$ どのくらい「期待から外れたか」すなわちサプライズを表現しているわけだ。F がサプ

ライズの上限を決めているので、$q(x)$ 最適化することで F を最小化するということは、最終的には最も驚きがない形に認識を合わせていくということを意味している。よって、$q(x)$ の最適なパラメータを ϕ^* とすれば、それはどんな ϕ かというと

$$\phi^* = \arg\min_{\phi} F$$

である。これは何をしていることになるのか？$q(x)$ は認識分布であった。これまでは、認識分布 $q(x)$ のパラメータ ϕ を変えることにより F を下げようとしたわけだ。認識を観測に合わせて最適化するわけなので、この操作は知覚に相当する。最小化を実行するには、F をパラメータで微分し、$\frac{\partial F}{\partial \phi}$ を定義する。この勾配の方向に ϕ が変化する更新則を考えれば、知覚のダイナミクスが定式化されたことになる。

他に F を最小化にする方法はあるのか。第1に、行動することを通じて、観測 o をより予測に合うようにサンプリングすることでも、F は下げられる。第2に、生成モデルを変更することだ。生成モデルのパラメータを θ とすると、F を下げるように θ を変更するわけだ。これは学習に相当する。この2つの F の下げ方については後述するが、単に受動的に観測基づいて認識を変容させるだけでなく、行動と学習を通じて自由エネルギーを下げることを能動的推論 (active inference)と呼ぶ。

ここまでの説明には、一点誤魔化しがあった。おさらいすると、$D_{KL}[q(x)\|p(x\,|\,o)] = F + \log p(o)$ (ただし $F = \sum_x q(x)\log\frac{q(x)}{p(x,o)}$) にカルバック・ライブラー情報量を分解できたというのがこれまでの流れなわけだ。この変形には $p(x\,|\,o) = p(x,o)/p(o)$ を使っている。しかし、左辺の $p(x\,|\,o)$ は o と x の真の関係、いわば真の事後分布を想定していたが、左辺の $p(x,o)$ はエージェントが保持している生成モデルである。よって、この等式が本来成立するのは、生成モデル $p(x,o)$ が厳密に真の同時確率分布 $p^*(x,o)$ と一致しているときである。生成モデルが実世界をまるで反映していなければ、いかに $q(x)$ を最適化しようと外界への予測は大外しするだろう。従って、上式はあくまで近似であることには留意する必要がある。ところが、ここが能動的推論の面白いところではあるのだが、現実には、生成モデルが世界を正しく反映していないときのほうが最適な行動を導くことがあるというシミュレーションも報告されている (Tschantz et al., 2020)。生成モデルは、エージェントが環境と相互作用する上で「役立てる」ためのモデルであって、世界を正確に写し取る鏡ではない。Ramstead et al. (2020) は、そのことから認知主義における表象主義を批判したりもしているのだが、それは自由エネルギー原理の哲学的な含意であるため、この辺でやめにしておく。

自由エネルギー F には、別の表し方がある。それを知っておくと、F の最小化の意味も捉えやすいだろう。

$$
\begin{aligned}
F &= \sum_x q(x) \log \frac{q(x)}{p(x, o)} \\
&= \sum_x q(x) \big(\log q(x) - \log p(o \mid x) - \log p(x) \big) \\
&= \sum_x q(x) \big(\log q(x) - p(x) \big) - \sum_x q(x) \log p(o \mid x) \\
&= D_{KL}[q(x) \| p(x)] - \mathbb{E}_{q(x)}[\log p(o \mid x)]
\end{aligned}
$$

最終行の第1項は認識分布 $q(x)$ と事前分布 $p(x)$ のカルバック・ライブラー情報量で、この量は「複雑さ (complexity)」と呼ばれている。第2項は、$q(x)$ の下での対数尤度で「正確さ (accuracy)」と呼ばれる。F を最小化しようとしたら、両者のトレードオフが生じる。というのは、$q(x)$ と $p(x)$ が近い方が D_{KL} は下がる。つまり、事前信念から変更がない方が第1項は小さくなる一方、$q(x)$ を尤度 $p(o \mid x)$ が高くなるように変更した方が第2項は下がっていくためである (符号反転している点に注意)。

ひとまず、ここまでで F を最小化すればエージェントが知覚、行動、学習を遂行することができるということがわかった。しかし、議論が抽象的で、具体的なモデルについては何も言っていなかった。実際に行動現象や神経活動を説明・予測するには、以上の原理を満たす形で「プロセス理論」あるいは「プロセスモデル」を導出する必要がある。Bogacz (2017) は知覚を例に実際に最適化を行う例を示している。

3. 期待自由エネルギー

行動を伴う自由エネルギーの最小化、つまりは能動的推論には、期待自由エネルギーという新たな量を導入する必要が出てくる。これがとても面倒臭い。

能動的推論では、何らかの行為に際して、その瞬間ではなく τ ステップ後の自由エネルギーを下げなければならないと考える。何らかの行為を生み出す制御状態を u としよう。u_t は、現在の制御状態である。自由エネルギーは観測 o を受け取ることで計算される量だが、この場合、制御 u_t の結果、τ ステップ後の観測 o_τ が生じる。この o_τ を受け取った際の自由エネルギーを下げたいわけだ。

$$
F_\tau = \sum_x q(x_\tau \mid u_t) \log \frac{q(x_\tau \mid u_t)}{p(x_\tau, o_\tau \mid u_t)}
$$

この時点では、x と o が τ ステップ後の値であるというのを強調した点と、各分布が制御 u_t で条件づけられている点以外、上で定義した自由エネルギーと見た目は何も変わっていない。x、o は将来の値 τ だが、制御状態は現在の値 u_t な点に注意しよう。現在の時刻 t の時点では o_τ は当然まだ観測されていない。通常の F の最小化の際は、観測 o が既に得られているが、o_τ は将来の値

なので、未知であるということだ。そこで能動的推論では、実際の o の観測値の代わりに o_τ の期待値を用いることにする。これはいわば、エージェントの将来の観測に関する信念として解釈できる (Tschantz et al., 2020)。モデルとしては、o_τ についての分布 $q(o_\tau|x_\tau)$ を仮定する。つまり、未観測の (将来の) 隠れ原因 x_τ に依存して観測 o_τ が生み出される過程をエージェントは信念として持っているということだ。また、$q(o_\tau|x_\tau)$ はパラメータ ϕ_τ を持つものとする。繰り返しになるが、将来の自由エネルギー F_τ は、未観測の o_τ に依存する。そこで、代わりに $q(o_\tau|x_\tau)$ で期待値を取って、「o_τ に関する信念」で代替する。すなわち、F_τ を次のように書き直す。

$$F_\tau = \sum_o q(o_\tau|x_\tau, u_t) \sum_x q(x_\tau|u_t) \log \frac{q(x_\tau|u_t)}{p(x_\tau, o_\tau|u_t)}$$

$q(o_\tau|x_\tau)$ はパラメータとして ϕ_τ を持っているとする。さらに、$q(o_\tau|x_\tau)q(x_\tau) = q(x_\tau, o_\tau)$ なので、上式は次のように変形できる。

$$F_\tau = \sum_{x,o} q(x_\tau, o_\tau|u_t) \log \frac{q(x_\tau|u_t)}{p(x_\tau, o_\tau|u_t)}$$

$$= \mathbb{E}_{q(x_\tau, o_\tau|u_t)}\left[\log \frac{q(x_\tau|u_t)}{p(x_\tau, o_\tau|u_t)}\right]$$

最後に、実際の観測ではなく将来の観測の期待値を使って計算される自由エネルギーを期待自由エネルギー G と書くことにする。

$$G = \sum_{x,o} q(x_\tau, o_\tau|u_t) \log \frac{q(x_\tau|u_t)}{p(x_\tau, o_\tau|u_t)} = \mathbb{E}_{q(x_\tau, o_\tau|u_t)}\left[\log \frac{q(x_\tau|u_t)}{p(x_\tau, o_\tau|u_t)}\right]$$

これで期待自由エネルギー G が定義できた。目下やったことをおさらいしよう。能動的推論では、制御状態 u_t すなわち行動によって、観測が変化する事態を考えるのだった。すると、考える自由エネルギーは現在の観測 o_t ではなく、u_t によって環境に働きかけた結果得られる将来の観測 o_τ に基づいて計算される必要がある。しかし、o_τ は当然現在の時刻 t の時点では未観測である。なので、そのままでは τ ステップ後の自由エネルギーは計算できない。そこで、実際の観測の代わりに、将来の観測についての信念で代替することにした。それは、$q(o_\tau|x_\tau)$ で期待値をとることに相当する。$q(o_\tau|x_\tau)$ はエージェントが持っている信念である。それにより、期待自由エネルギー G を定義することができた。能動的推論では、この G を最小化するようにエージェントは行動すると考える。つまり、G が小さくなる方向に u_t をとるということだ。

期待自由エネルギーは、少し変形と近似をすることで直観的な意味合いを持つようになる。それを見ていこう。まず、$p(x_\tau, o_\tau | u_t) = p(x_\tau | o_\tau, u_t)p(o_\tau)$ である。また、$p(o_\tau)$ は観測について事前信念であると考えられるが、能動的推論では $p(o_\tau)$ を「選好」を表現しているとみなす。つまり、エージェントがより予期している事象が好まれる。というよりも、能動的推論では、選好される事象に高い事前確率を付与していると仮定されるわけだ。

$$G = \mathbb{E}_{q(x_\tau, o_\tau | u_t)}\left[\log \frac{q(x_\tau | u_t)}{p(x_\tau, o_\tau | u_t)}\right]$$

$$= \mathbb{E}_{q(x_\tau, o_\tau | u_t)}\left[\log \frac{q(x_\tau | u_t)}{p(x_\tau | o_\tau, u_t)p(o_\tau)}\right]$$
$$= \mathbb{E}_{q(x_\tau, o_\tau | u_t)}\left[\log \frac{q(x_\tau | u_t)}{p(x_\tau | o_\tau, u_t)} - \log p(o_\tau)\right]$$

$$= \mathbb{E}_{q(x_\tau, o_\tau | u_t)}\left[\log \frac{q(x_\tau | u_t)}{p(x_\tau | o_\tau, u_t)}\right] - \mathbb{E}_{q(x_\tau, o_\tau | u_t)}\left[\log p(o_\tau)\right]$$

しかし、$p(x_\tau | o_\tau, u_t)$ は事後分布であるため、(o_τ が未観測であることを抜きにしても) 直接求めることができない。そこで、$q(x_\tau | o_\tau, u_t) \approx p(x_\tau | o_\tau, u_t)$ であるとする。認識分布が十分に事後分布を近似できていると仮定するわけだ。すると、期待自由エネルギーは次のように書き換えられる。

$$G \approx \mathbb{E}_{q(x_\tau, o_\tau | u_t)}\left[\log \frac{q(x_\tau | u_t)}{q(x_\tau | o_\tau, u_t)}\right] - \mathbb{E}_{q(x_\tau, o_\tau | u_t)}\left[\log p(o_\tau)\right]$$

右辺第1項は、負の認識的価値 (あるいは内在的価値)、第2項は道具的価値 (あるいはプラグマティック価値、外在的価値、目的志向的価値) と呼ばれる。両者とも$q(x_\tau, o_\tau | u_t)$ で期待値を取っているものの、期待値の中身を見てみると意味合いがわかる。まず第1項の負の認識的価値は、$q(x_\tau | u_t)$ と $q(x_\tau | o_\tau, u_t)$ が似ているほど、低くなる値である。違いは、o_τ で条件づけているかだけである。つまり、この項は将来の観測 (の信念) の情報を入れることで、隠れ状態 x_τ についての信念がどれくらい変化するかを表している。$\log \frac{q(x_\tau | u_t)}{q(x_\tau | o_\tau, u_t)} = \log q(x_\tau | u_t) - \log q(x_\tau | o_\tau, u_t)$ であるため、o_τ で条件づけている、あるいは条件づけていない認識分布の間の引き算になっている。平易に言えば、環境について知ることで不確実性が減れば、この項は小さくなる。

第2項の道具的価値は、$q(x_\tau, o_\tau | u_t)$ の下での選好の期待値になっている。つまり、この量はそのままエージェントの選好を表現している。よって、第1項は「環境についてどれくらい知識が増えるか」を表す一方、第2項は選好に基づいて価値を決める項になっている。故に、G を下げるには2つの方法がある。最初に、信念 $q(x_\tau | u_t)$ を $q(x_\tau | o_\tau, u_t)$ に合うように変更することである。 つま

り、o を得なくても $q(x_\tau|u_t)$ だけで環境について十分に予測ができてしまう状態に学習していくということである。これは認識的価値を高めることに相当する。次に、選好に合うように観測を得ることが挙げられる。選好は $p(o_\tau)$ で表現されていると仮定したので、これは道具的価値が高い制御 u_t を実行すればよい。G を最小化するには、認識的価値と道具的価値両者のバランスを取る必要が出てくる。換言すれば、G を最小化によって、自然と探索と搾取のトレードオフを定式化していることになる。

認識的価値がややわかりづらいと思う人もいるかもしれない。認識的価値を少し変形させたら、その意味合いがクリアになるだろう。上式では、負の認識的価値を $\mathbb{E}_{q(x_\tau, o_\tau|u_t)}\left[\log \frac{q(x_\tau|u_t)}{q(x_\tau|o_\tau, u_t)}\right]$ と定義した。正の認識的価値はこの符号反転である。それを変形していくとしよう。

$$-\mathbb{E}_{q(x_\tau, o_\tau|u_t)}\left[\log \frac{q(x_\tau|u_t)}{q(x_\tau|o_\tau, u_t)}\right] = \mathbb{E}_{q(x_\tau, o_\tau|u_t)}\left[\log \frac{q(x_\tau|o_\tau, u_t)}{q(x_\tau|u_t)}\right]$$

$q(x_\tau, o_\tau|u_t) = q(o_\tau|u_t)q(x_\tau|o_\tau, u_t)$ を使えば

$$= \mathbb{E}_{q(o_\tau|u_t)q(x_\tau|o_\tau, u_t)}\left[\log \frac{q(x_\tau|o_\tau, u_t)}{q(x_\tau|u_t)}\right]$$

$$= \sum_{x,o} q(o_\tau|u_t)q(x_\tau|o_\tau, u_t)\log \frac{q(x_\tau|o_\tau, u_t)}{q(x_\tau|u_t)}$$

$$= \sum_{o} q(o_\tau|u_t)\sum_{x} q(x_\tau|o_\tau, u_t)\log \frac{q(x_\tau|o_\tau, u_t)}{q(x_\tau|u_t)}$$

$$= \mathbb{E}_{q(o_\tau|u_t)}[D_{KL}[q(x_\tau|o_\tau, u_t)\|q(x_\tau|u_t)]]$$

最後の変形には、カルバック・ライブラー情報量の定義そのものを使っている。$q(x_\tau|o_\tau, u_t)$ と $q(x_\tau|u_t)$ はともに認識 (事後) 分布であるが、違いは o_τ で条件づけているかの部分にある。つまり、期待される観測 o_τ によって将来の隠れ状態 x_τ の分布に変化が起きるかを、認識的価値は表現している。変わるなら価値が高い。つまり、認識的価値は、観測を得ることで世界について知ることが増えそうならば、その行動には価値があるという意味を持つ。故に、認識的価値によって、環境の不確定性を下げることがエージェントにとって価値を持つという解釈が成立する。エージェントは、環境から得られる情報が高い選択を取ると言ってもよい。負の認識的価値はこの符号反転なので、認識的価値が高い行動を取れば G は結果的に下がるという寸法だ。また、$D_{KL}[q(x_\tau|o_\tau, u_t)\|q(x_\tau|u_t)$ というカルバック・ライブラー情報量にはベイズサプライズという名前がついている。

なお、FristonをはじめとしたPOMDP状況での能動的推論では、ここまでで見てきたように自由エネルギーの期待値を取るという面倒なことをする。結果、プロセスモデルを実際に導出する際も、その更新則がなかなか難解なものになる (Smith et al., 2021; Tschantz et al., 2020)。一方、

Bogacz (2020) はそこを避け、正規分布だけで書いたシンプルな形で、動物の反応率を予測するプロセスモデルを提案している。

Reference

Bogacz, R. (2017). A tutorial on the free-energy framework for modelling perception and learning. Journal of mathematical psychology, 76, 198-211.

Bogacz, R. (2020). Dopamine role in learning and action inference. Elife, 9, e53262.

Friston, K. (2005). A theory of cortical responses. Philosophical transactions of the Royal Society B: Biological sciences, 360(1456), 815-836.

Friston, K. (2010). The free-energy principle: a unified brain theory?. Nature reviews neuroscience, 11(2), 127-138.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: a process theory. Neural computation, 29(1), 1-49.

Ramstead, M. J. D., Badcock, P. B., & Friston, K. J. (2018). Answering Schrödinger's question: A free-energy formulation. Physics of life reviews, 24, 1-16.

Ramstead, M. J., Kirchhoff, M. D., & Friston, K. J. (2020). A tale of two densities: Active inference is enactive inference. Adaptive Behavior, 28(4), 225-239.

島崎秀昭, 吉田正俊, 田口茂, 磯村拓哉, 田中琢真, 大羽成征, & 乾敏郎. (2018). 特集 「自由エネルギー原理入門」. 日本神経回路学会誌, 25(3), 51-52.

Smith, R., Friston, K., & Whyte, C. (2021). A Step-by-Step Tutorial on Active Inference and its Application to Empirical Data, psyarxiv, https://psyarxiv.com/b4jm6/

Tschantz, A., Seth, A. K., & Buckley, C. L. (2020). Learning action-oriented models through active inference. PLoS computational biology, 16(4), e1007805.

吉田正俊, & 田口茂. (2018). 自由エネルギー原理と視覚的意識. 日本神経回路学会誌, 25(3), 53-70.

Appendix

A1. ベイズの定理

2つの確率変数 x と y があるとする。それぞれの確率分布は $p(x)$、$p(y)$、その同時分布を $p(x, y)$ と書く。条件付き確率分布は$p(x|y) = \frac{p(x, y)}{p(y)}$ と定義される。すると

$$p(x, y) = p(x|y)p(y)$$

かつ

$$p(x, y) = p(y|x)p(x)$$

が成り立つ。そこで両者を等式で結ぶと

$$p(x|y)p(y) = p(y|x)p(x)$$

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

これがベイズの定理である。

A2. 期待値の定義

$p(x)$ を x の確率分布とする。x の期待値は次のように定義される。

$$\mathbb{E}[p(x)] = \sum_x xp(x)$$

ようは期待値とは、確率分布 $p(x)$ でありうる x 全てを重み付けた平均である。

A3. Shannonの情報量、エントロピー、カルバック・ライブラー情報量

$p(x)$ を x の確率分布とする。x の情報量は次のように定義される。

$$I[x] = -\log p(x)$$

対数の底が2のとき、単位はビット (bit)、ネイピア数 e のとき、ナット (nat) と呼ばれる。エントロピーとは、情報量の期待値である。よって、次のように定義される。

$$\mathbb{E}[I_x] = -\sum_x p(x)\log p(x)$$

カルバック・ライブラー情報量は、二つの分布間の類似性を表す指標である。