自由エネルギー原理2
期待自由エネルギー

https://medium.com/@solopchuk/tutorial-on-active-inference-30edcf50f5dc

のまとめ

公立小松大学

藤田 一寿

スライドに間違いがあるかもしれないし内容が古いので，研究で使う際は必ず論文(Smith et al., 2022; Sajid et al., 2021など)をチェックすること！！

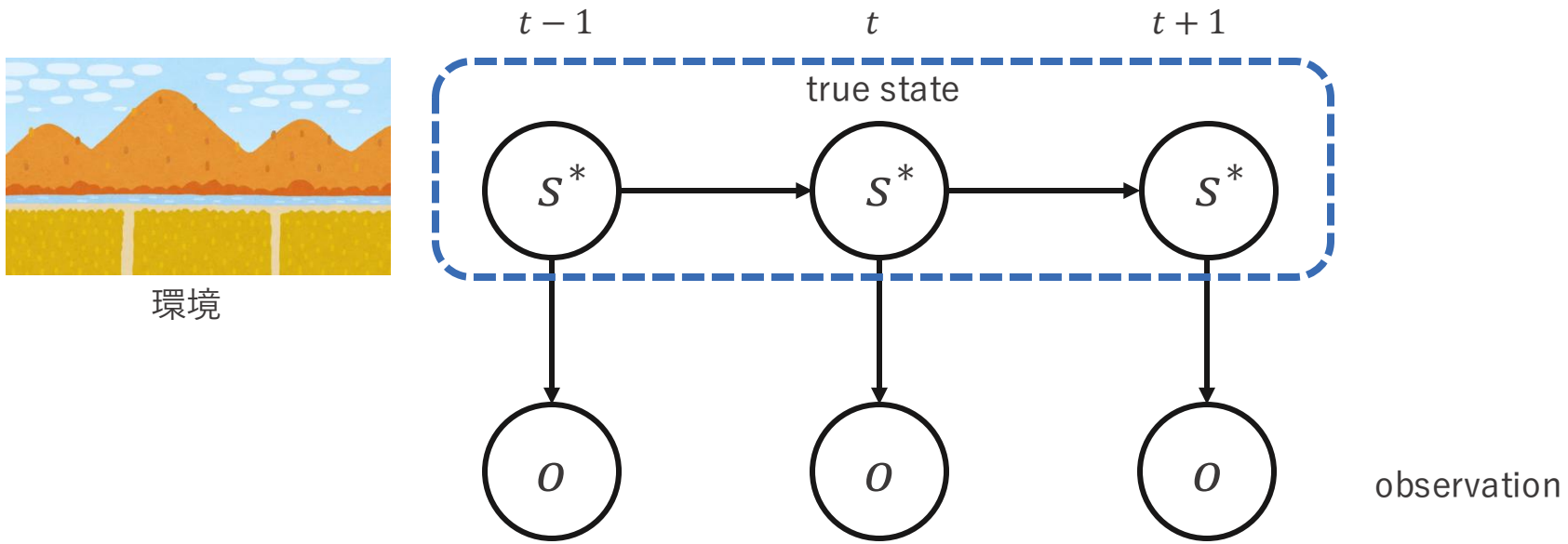
途中式があるので，論文を読むときの参考になるかも．

Ver. 20241030

方策と行動

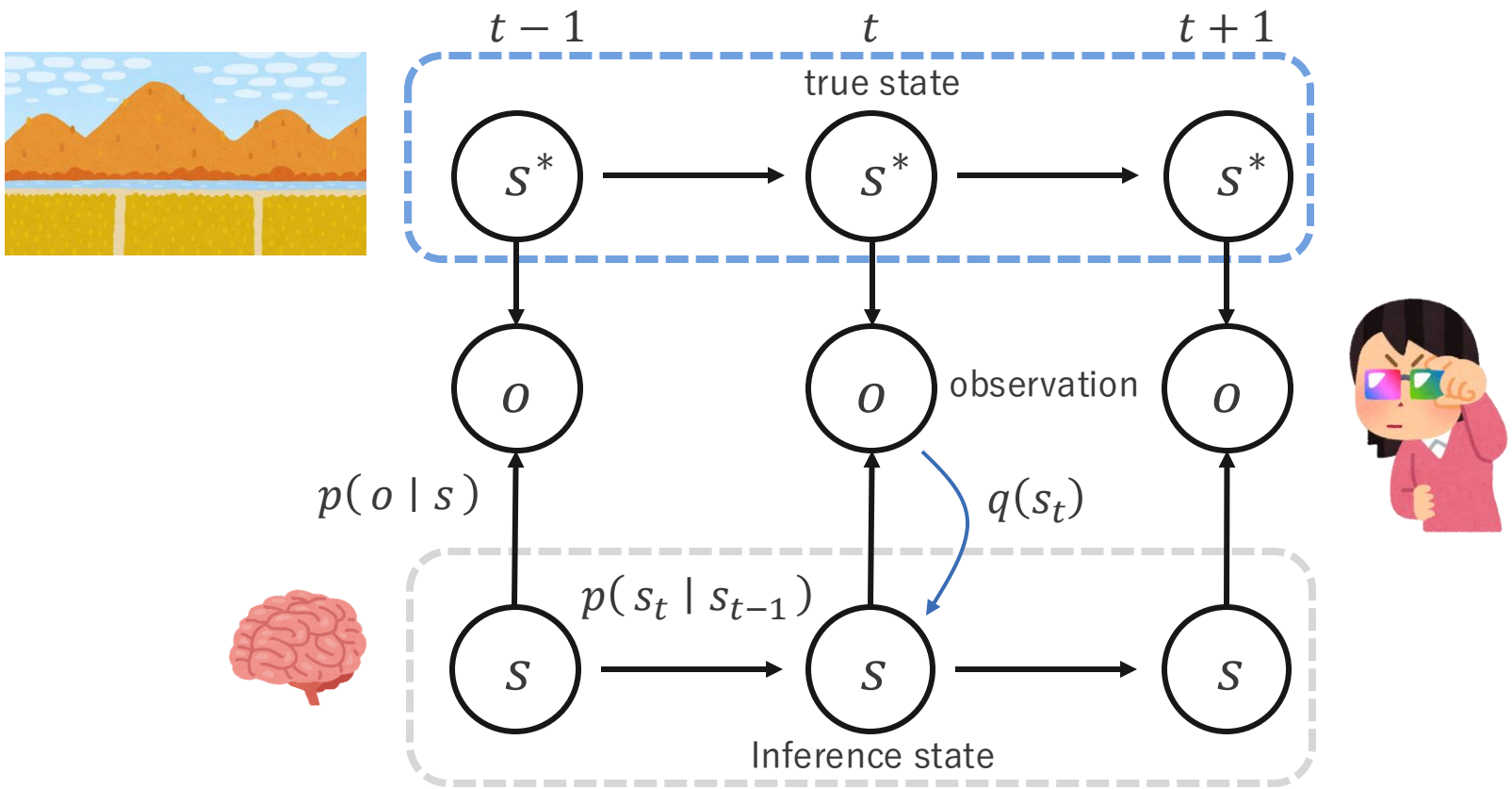
時系列で考える

- 環境の状態は時間とともに変化する.
- 状態は直前の状態に依存するとする.
- それぞれの状態から，それに対応した観測が生まれる.



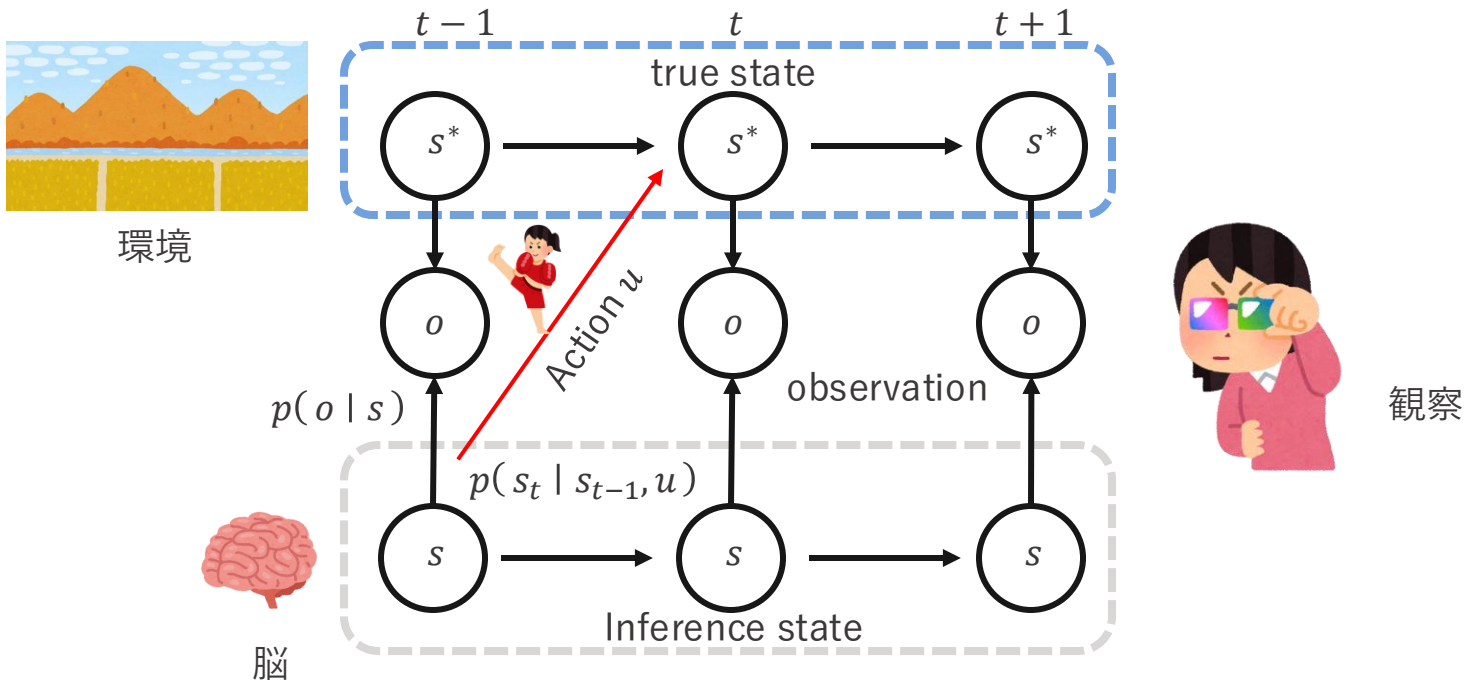
時系列で考える

- Agentは生成モデル$p(o,s)$の学習とそれぞれの時間で事後分布$q(s)$の近似を得ることにより，真の生成過程$p(o,s^*)$をモデル化しようとする．
- 簡単な場合では，自由エネルギーを減らすようにパラメタを変えることで探すことができる（自由エネルギーのスライド参照）．



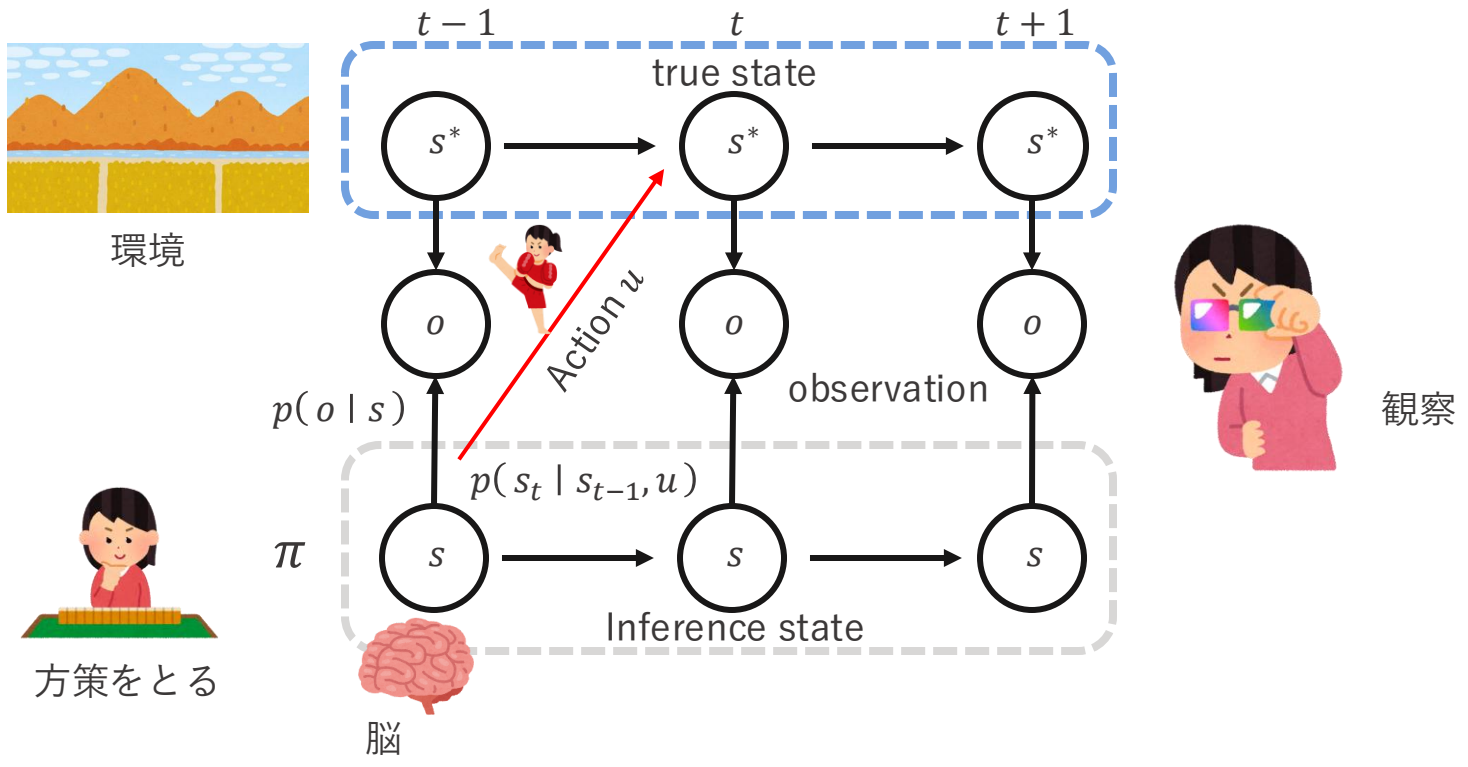
行動してみる

- 先の例は，環境の状態を受動的に観測するだけだった．
- Agentが行動をする場合，その行動により状態が変わる．
- つまり，行動が直接環境に影響を与え，異なる行動は異なる未来を導くことになる．



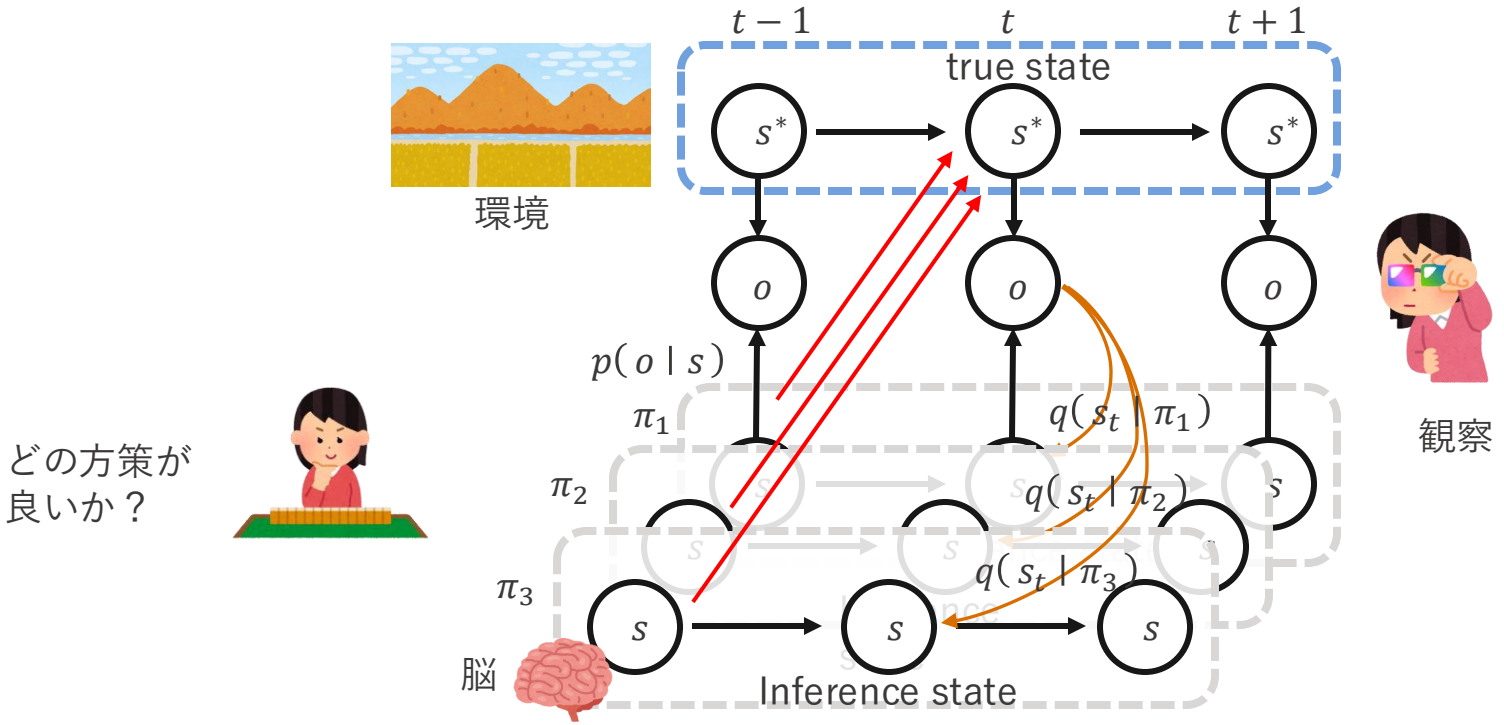
我々はどのような行動を取ればよいのか？

- Agentは当然それぞれの時間で良い行動を選びたい.
- 一方で，Agentは行動直後の結果のみを考えて行動しているのではなく，時間的に離れた目標に向けて一連の行動をしている.
- この一連の行動のルールを方策（policy）πという.



方策

- Agentが取ることの出来る方策はたくさんある.
- Active inferenceでは，それらすべてを考える.
- だから，Agentはすべての可能な方策πに対し，$p(s \mid o)$を$q(s)$で近似し推論する.
- **将来の自由エネルギーを最小化する方策が優先される.**



強化学習では将来得られる報酬が多い行動が優先される.
強化学習では，方策は$p(u \mid s)$.

期待自由エネルギー

期待自由エネルギー

- 将来の自由エネルギーを最小化するためには，　将来の自由エネルギーを知る必要がある．
- 将来どれほどの自由エネルギーになるかを知るためには自由エネルギーの期待値を取る必要がある．
- 将来の自由エネルギーはAgentがとる方策にも依存する．

期待自由エネルギー

- 自由エネルギーの式を，方策πを考慮したものに書き換える.
- $\sum_s q(s) \log \frac{q(s)}{p(o,s)} \rightarrow \sum_s q(s_t \mid \pi) \log \frac{q(s_t|\pi)}{p(o_t,s_t|\pi)}$
- 更に$p(o_t)$について期待値をとる.
- $G = \sum_o p(o_t \mid s_t) \sum_s q(s_t \mid \pi) \log \frac{q(s_t|\pi)}{p(o_t,s_t|\pi)}$
- ここではo_tとs_tの関係はpolicyによらないとしている.
- さらに式変形すると
- $G = \sum_{o,s} q(s_t \mid \pi) p(o_t \mid s_t) \log \frac{q(s_t|\pi)}{p(o_t,s_t|\pi)}$
- $= \sum_{o,s} q(s_t \mid \pi) p(o_t \mid s_t) \log \frac{q(s_t|\pi)}{p(s_t|o_t,\pi)p(o_t)}$

期待自由エネルギー

更に式変形する

- $G = \sum_{o,s} q(s_t \mid \pi) p(o_t \mid s_t) \log \frac{q(s_t|\pi)}{p(s_t|o_t,\pi)p(o_t)}$
- $= \sum_{o,s} q(s_t \mid \pi) p(o_t \mid s_t) \log \frac{q(s_t|\pi)}{p(s_t|o_t,\pi)} - \sum_{o,s} q(s_t \mid \pi) p(o_t \mid s_t) \log p(o_t)$
- $= -\sum_{o,s} q(s_t \mid \pi) p(o_t \mid s_t) \log \frac{p(s_t|o_t,\pi)}{q(s_t|\pi)} - \sum_{o,s} q(s_t \mid \pi) p(o_t \mid s_t) \log p(o_t)$

- $\sum_{o,s} q(s_t \mid \pi) p(o_t \mid s_t) \log \frac{p(s_t|o_t,\pi)}{q(s_t|\pi)}$をepistemic valueという.

さらに式変形する

- $G = \sum_{o,s} q(s_t \mid \pi) p(o_t \mid s_t) \log \frac{q(o_t|\pi)}{p(o_t|s_t,\pi)} - \sum_{o,s} q(s_t \mid \pi) p(o_t \mid s_t) \log p(o_t)$
- $= \sum_{o,s} q(s_t \mid \pi) p(o_t \mid s_t) \log \frac{q(o_t|\pi)}{p(o_t|s_t,\pi)p(o_t)}$
- $= \sum_{o,s} q(s_t \mid \pi) p(o_t \mid s_t) \log \frac{q(o_t|\pi)}{p(o_t)} - \sum_{o,s} q(s_t \mid \pi) p(o_t \mid s_t) \log p(o_t \mid s_t, \pi)$
- 近似が十分正確だとすれば$q(o_t \mid s_t) = p(o_t \mid s_t)$と見なせるので
- $G = \sum_{o,s} q(s_t \mid \pi) q(o_t \mid s_t) \log \frac{q(o_t|\pi)}{p(o_t)} - \sum_{o,s} q(s_t \mid \pi) p(o_t \mid s_t) \log p(o_t \mid s_t, \pi)$
- $= \sum_{o,s} q(o_t, s_t \mid \pi) \log \frac{q(o_t|\pi)}{p(o_t)} - \sum_{s} q(s_t \mid \pi) \sum_{o} p(o_t \mid s_t) \log p(o_t \mid s_t, \pi)$
- $= \sum_{o} q(o_t, \mid \pi) \log \frac{q(o_t|\pi)}{p(o_t)} - \sum_{s} q(s_t \mid \pi) \sum_{o} p(o_t \mid s_t) \log p(\, o_t \mid s_t \,)$
- $= KL\big(q(o_t \mid \pi) || p(o_t)\big) + \sum_{s} q(s_t \mid \pi) H[p(\, o_t \mid s_t \,)]$

o_sとs_tの関係はpolicyによらない

$$H[p(\, o_t \mid s_t \,)] = -\sum_{o} p(o_t \mid s_t) \log p(\, o_t \mid s_t \,)$$

最終的な期待自由エネルギーの式

- $G = KL\big(q(o_t \mid \pi)||p(o_t)\big) + \sum_s q(s_t \mid \pi) H[p(o_t \mid s_t)]$

  Expected cost　　Expected ambiguity

- Expected costは，方策πの下での予想される観測$q(o_t \mid \pi)$と**prior preferences** $p(o_t)$の2つの分布の間のKLダイバージェンスである．つまり，期待自由エネルギーを最小化すると，Agentが望む観測をもたらすような方策を好むことになる．
- Expected Ambiguityは，$p(o \mid s)$のエントロピーの期待値である．つまり，状態と観測値間のマッピング$p(o|s)$がどれだけ不確実であるかを定量化している．

Prior prefernces
乾の訳では事前の選好とされていた．Agentが好む観測の分布を意味する．AgentはPrior preferencesを目指し行動する．
Smit et al.の論文ではPrior preference distributionを$p(o \mid C)$と表現する．変数Cは，エージェントのpreferencesを表すとされている．
Parr et al, 2022によればActive inferenceの論文ではCがしばしば省略されるらしい．

Estimetic valueの考察

Epistemic valueの変形

- $\sum_{o,s} q(s_t \mid \pi) p(o_t \mid s_t) \log \frac{p(s_t|o_t,\pi)}{q(s_t|\pi)}$
- $\log \frac{p(s_t|o_t,\pi)}{q(s_t|\pi)} = \log \frac{p(s_t|o_t,\pi) q(o_t|\pi)}{q(s_t|\pi) q(o_t|\pi)}$
- 推定が正確だとすれば$q(o_t \mid \pi) = p(o_t \mid \pi)$となるから
- $\log \frac{p(s_t|o_t,\pi) q(o_t|\pi)}{q(s_t|\pi) q(o_t|\pi)} = \log \frac{p(o_t,s_t|\pi)}{q(s_t|\pi) q(o_t|\pi)} = \log \frac{p(o_t|s_t,\pi) q(s_t|\pi)}{q(s_t|\pi) q(o_t|\pi)} = \log \frac{p(o_t|s_t,\pi)}{q(o_t|\pi)}$
- よって
- $\sum_{o,s} q(s_t \mid \pi) p(o_t \mid s_t) \log \frac{p(s_t|o_t,\pi)}{q(s_t|\pi)} = \sum_{o,s} q(s_t \mid \pi) p(o_t \mid s_t) \log \frac{p(o_t|s_t,\pi)}{q(o_t|\pi)}$

Epistemic valueは相互情報量

- $MI(a,b) = \sum_{ab} p(a,b) \log \frac{p(a,b)}{p(a)p(b)}$
- $= \sum_{ab} p(a \mid b)p(b) \log \frac{p(a|b)p(b)}{p(a)p(b)} = \sum_{ab} p(a \mid b)p(b) \log \frac{p(a|b)}{p(a)} = H[p(a)] - H[p(a \mid b)] = H[p(b)] - H[p(b \mid a)]$
- $MI(o,s) = \sum_{o,s} p(o_t \mid s_t)q(s_t \mid \pi) \log \frac{p(o_t|s_t,\pi)}{q(o_t|\pi)}$

Epistemic value

おまけ

$$MI(a,b) = \sum_{ab} p(a \mid b)p(b) \log \frac{p(a \mid b)p(b)}{p(a)p(b)} = \sum_{ab} p(a \mid b)p(b) \log \frac{p(a \mid b)}{p(a)}$$

$$= \sum_{ab} p(a \mid b)p(b) \log p(a \mid b) - \sum_{ab} p(a \mid b)p(b) \log p(a)$$

$$= \sum_{ab} p(a \mid b)p(b) \log p(a \mid b) - \sum_{a} p(a) \log p(a)$$

$$= H[p(a)] - H[p(a \mid b)] = H[p(b)] - H[p(b \mid a)]$$

Epistemic valueの解釈

- $MI(o,s) = \sum_{o,s} q(s_t \mid \pi) p(o_t \mid s_t) \log \frac{p(o_t|s_t,\pi)}{q(o_t|\pi)}$
- $= H[q(s_t \mid \pi)] - H[p(s_t \mid o_t)]$

- Agentが非常に確信している場合，$H[q(s_t \mid \pi)]$ は小さく，これ以上学ぶことは何もないので，Epistemic value（認識価値）は低くなる.
  - 確信していれば，方策πを選んだときに起こることが推測できるため，エントロピーが小さくなる.
  - 例：方策πをとったとき，必ず状態sになると確信していれば，$H[q(s_t \mid \pi)]$は0となる.
- 確信が持てない場合，$H[q(s_t \mid \pi)]$が高い.
  - 確信が持てていないため，どの状態になるか分からない.
  - 結果，Epistemic valueは高くなる .
  - 例：方策πをとったとき，どの状態になるか分からず，Agentがすべての状態が当確率に現れると思っていれば，$H[q(s_t \mid \pi)]$は最大値を取る.

エントロピーが最大，最小となる条件を確認しよう.

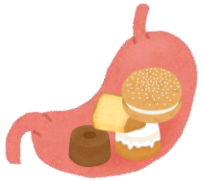
具体例で見るActive inferenceと期待自由エネルギー：準備

空腹かどうか

- お腹の空き具合と食べることを考える.
- 胃の中の状態sは，満杯1とカラ2の2種類である.
- 観測oは，満腹1と空腹2の2種類である.
- 生成モデル$p(o,s)$のパラメタは既知であるとする.

胃の中の状態s

1: 満杯



2: カラ

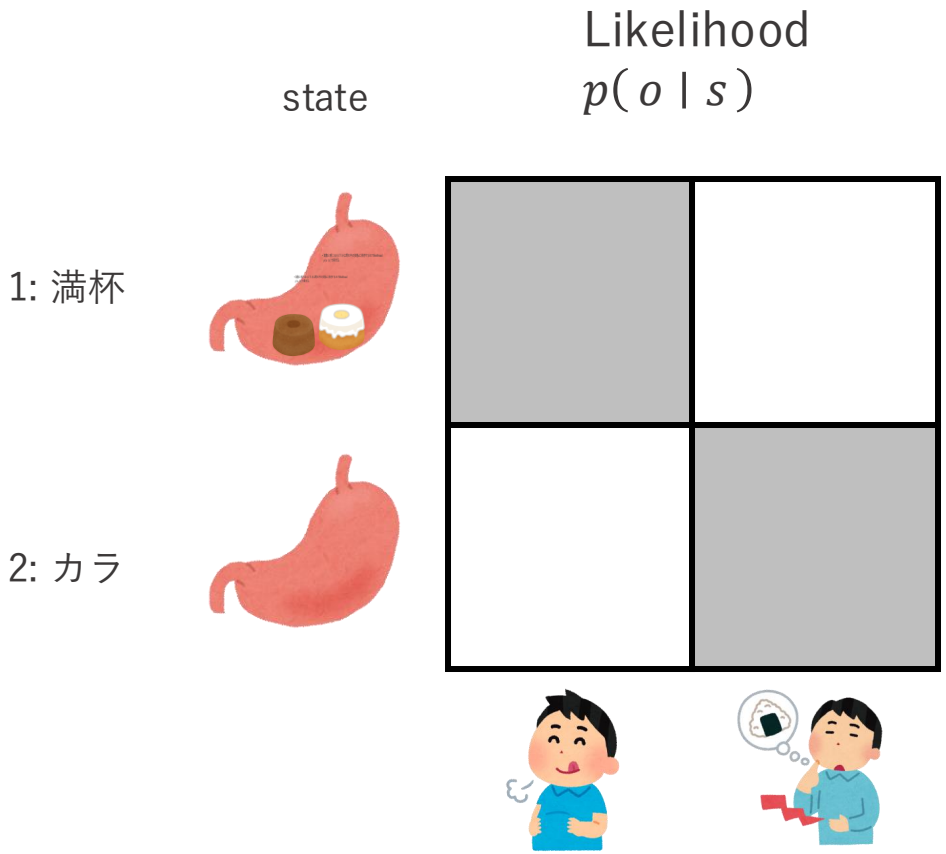
満腹かどうかo

1: 満腹

2: 空腹

満腹感と胃の状態

- 満腹と感じるかどうかは胃の中の状態sに依存するのでlikelihood $p(o \mid s)$で表せる.

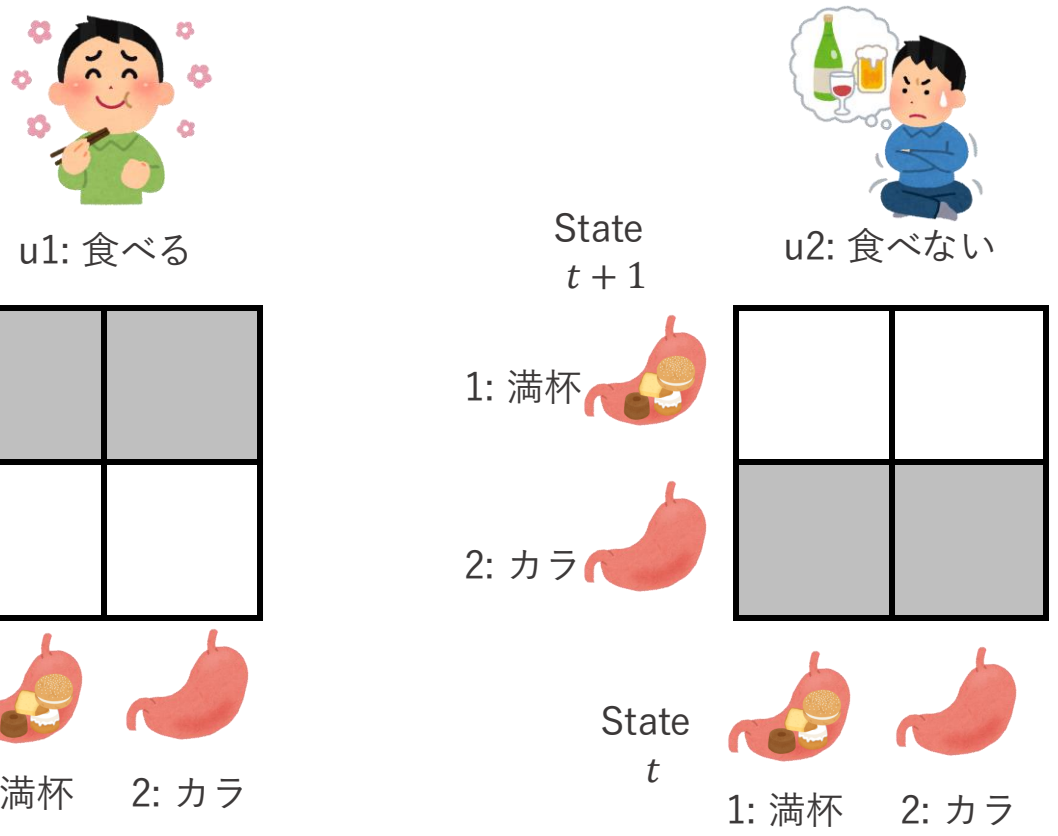


当然，胃が満杯だと満腹だし，カラだと空腹になる.

食べるかどうか

- 食べるかどうかはuで表す.
- 状態s_tは以前の状態と行動に依存するので$p(s_t \mid s_{t-1}, u)$と表せる.

Transition $p(s_t \mid s_{t-1}, u)$



食べれば胃は満杯になり，食べなければからになる.

空腹具合

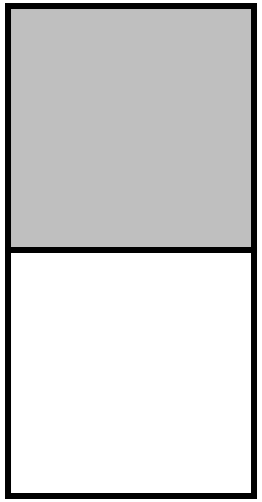
- Agentはprior preferences $p(o)$を持つ.
- Agentは空腹でないことを好むから，満腹が観測されることを好む.
- 観測に対する好みを確率 $p(o)$で表す.

Prior preferences

$p(o)$

1: 満腹

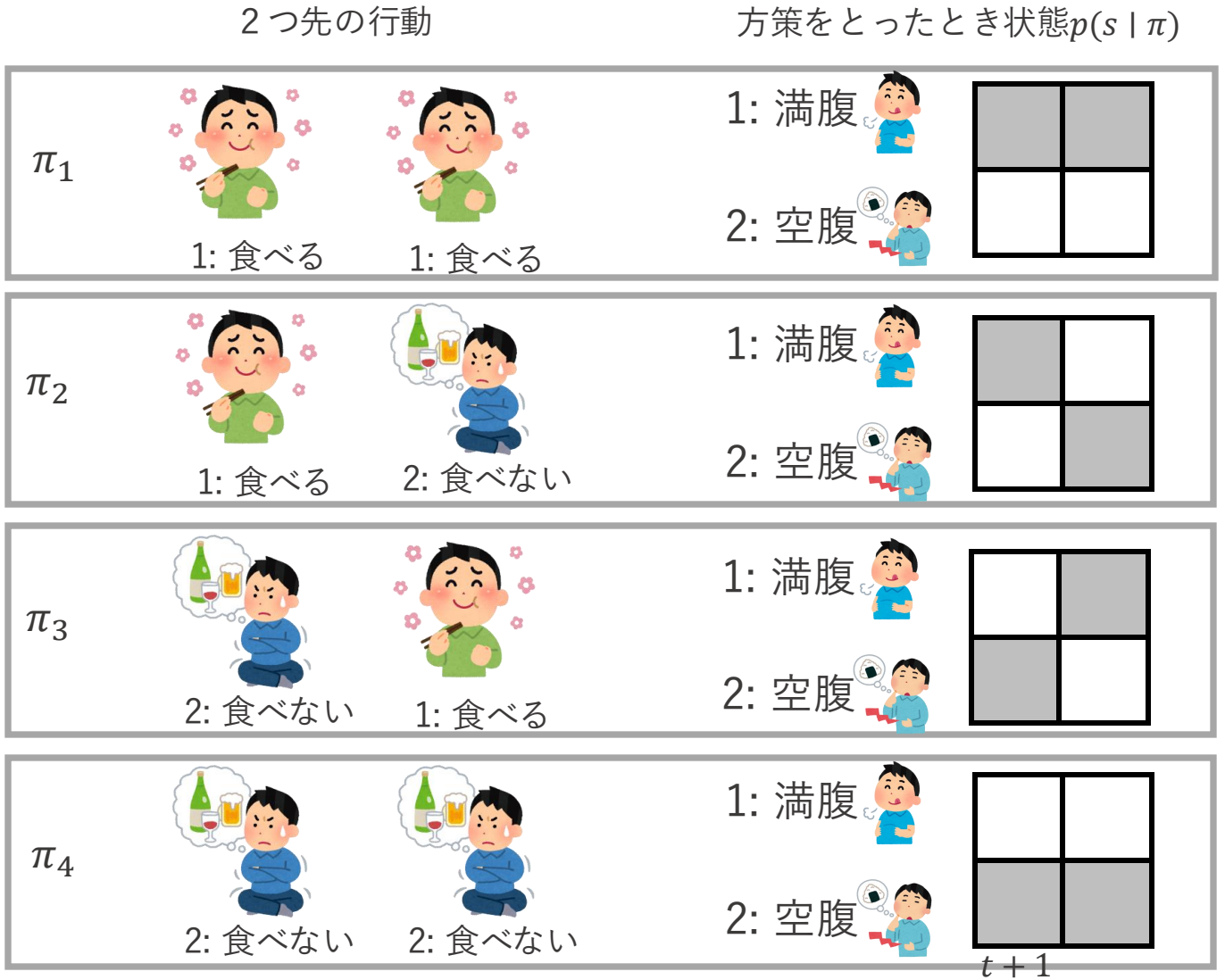
2: 空腹



四角は確率を表す．濃いほうが確率が高い．
空腹より満腹の方を好むので満腹のほうがprior preferencesが高い．

方策

- ２つ先の未来までの行動が方策で決定されるとすると，policyは次の4種類になる．
- π_1:食べる，食べる
- π_2:食べる，食べない
- π_3:食べない，食べる
- π_4:食べない，食べない



期待自由エネルギーの計算 -KLダイバージェンス-

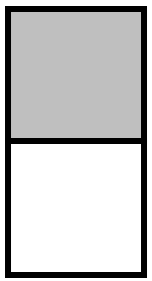
- Agentは状態と観測の関係$p(o \mid s)$を知っているから，各方策の予測した(predicted)観測$q(o \mid \pi)$を推定する(estimate)ことができる．
  - 方策πを決める→行動uする→状態sが変わる→観測oを得る，という流れだから方策さえ決まれば得られる観測がどうなるか推定できる．
- よって，各ポリシーの期待自由エネルギーのKL 項を計算できる．

$$G = \underline{KL\big(q(o_t \mid \pi) || p(o_t)\big)} + \sum_s q(s_t \mid \pi) H[p(o_t \mid s_t)]$$

Desired observation
$p(o)$

1: 満腹

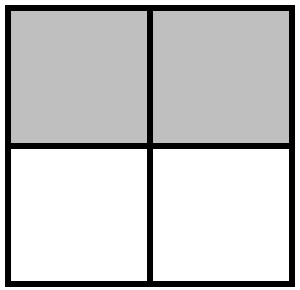
2: 空腹



Predicted observation
$q(o \mid \pi)$

1: 満腹

2: 空腹

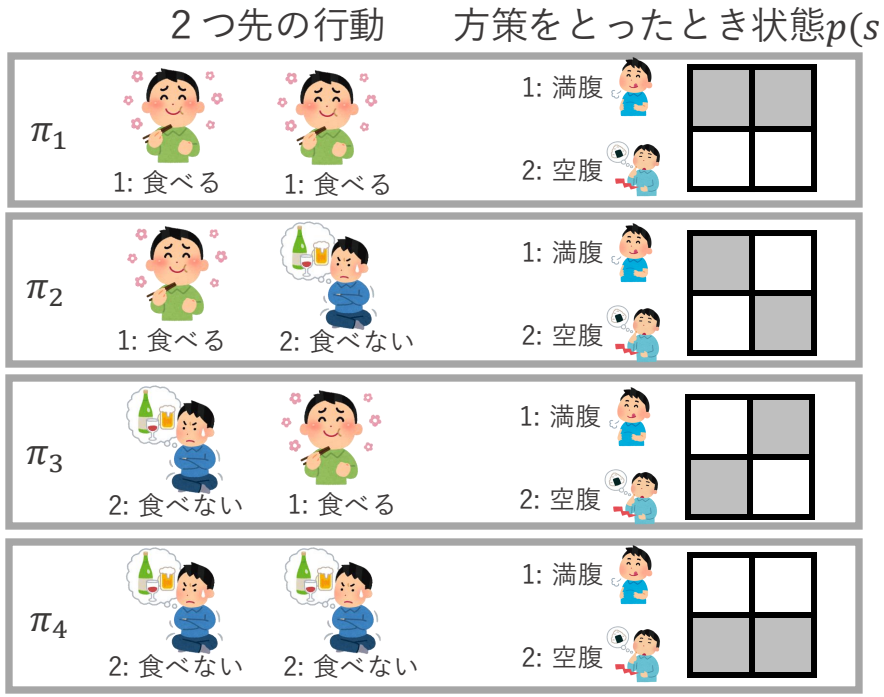


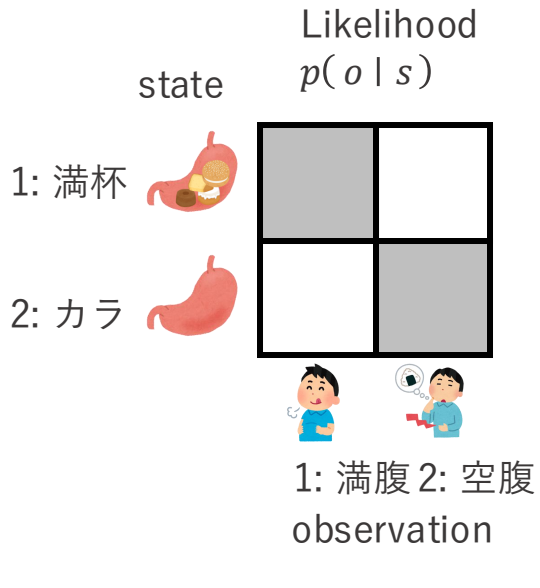
$p(o)$と$q(o \mid \pi)$のKLダイバージェンスが小さければ小さいほど，Agentの希望する結果を得られる可能性が高い．

期待自由エネルギーの計算 -ambiduity-

- 方策πが決まれば，どのような状態になるか推定できる．
- 状態sが決まれば，何が観測されるか推定できる．
- よって，$p(o \mid s)$ に依存するambiguity項も評価できる．

$$G = KL\big(q(o_t \mid \pi) || p(o_t)\big) + \sum_s q(s_t \mid \pi) H[p(o_t \mid s_t)]$$



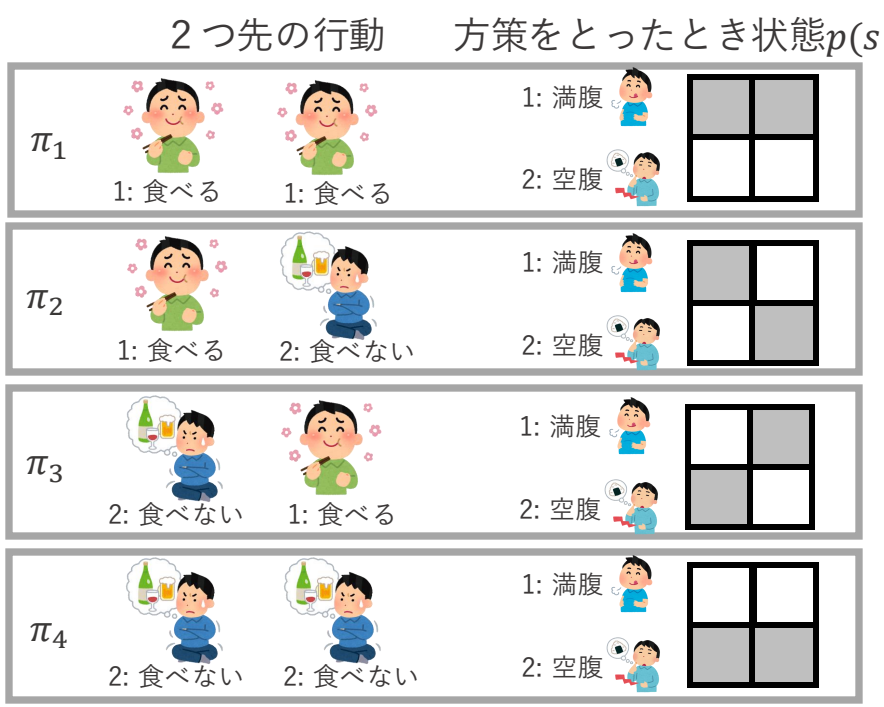


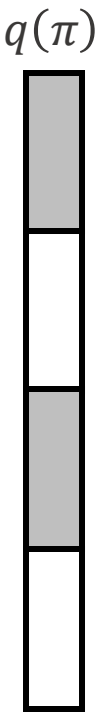
具体例で見るActive inferenceと期待自由エネルギー：次の行動を決める

どのようにして次の行動を決めるのか

- まず，将来の時間ステップで期待自由エネルギーを合計する．
- それを方策πに対する確率分布$q(\pi)$に変換する．
  - その確率は自由エネルギーが小さいほど高い．

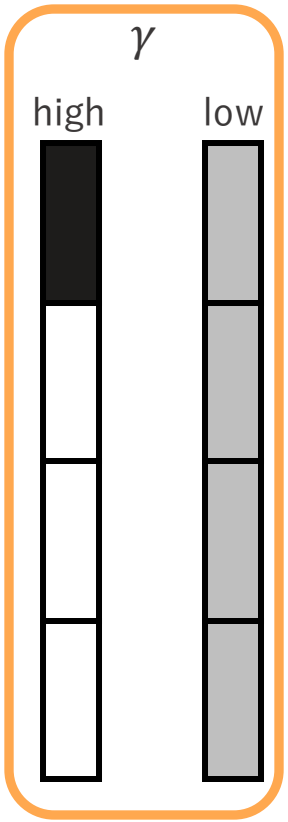
期待自由エネルギー $G = KL\big(q(o_t \mid \pi)||p(o_t)\big) + \sum_s q(s_t \mid \pi) H[p(o_t \mid s_t)]$





Precisionγをかける．

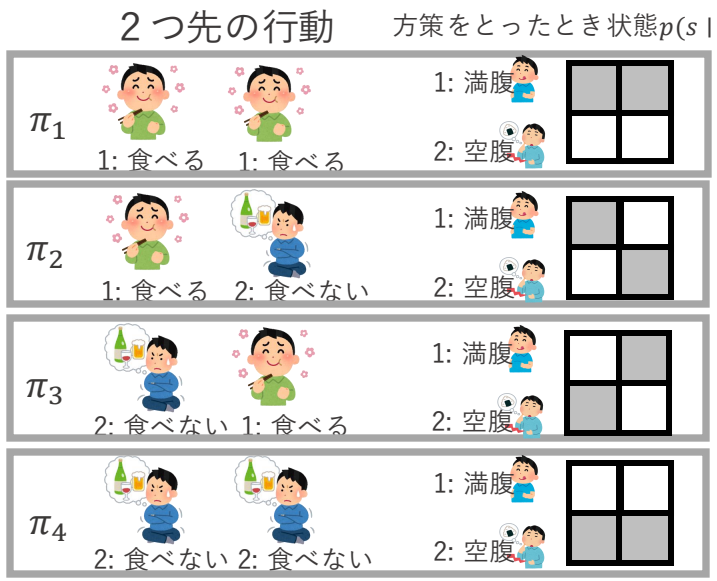
Softmax関数σで規格化する．
$q(\pi) = \sigma(-\gamma G(\pi))$

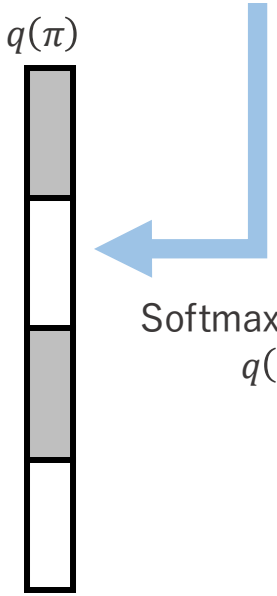


精度

- この変換の際に，自由エネルギーは精度γによって重み付けされる．
  - γは方策に対する信念(belief)をどれほど確信しているかを表す．
  - 精度を極端に変えることによって，agentの信念は一つの方策に集約されたり，一様に広がったりする．
  - これは探索と利用を決める上で重要である．良い方策を持っていると確信するほど(すなわち，精度が高いほど)探索は少なくなり，その逆もまた然りである．

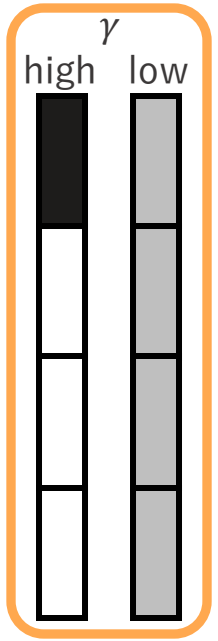
期待自由エネルギー　$G = KL\left(q(o_t \mid \pi)||p(o_t)\right) + \sum_s q(s_t \mid \pi)H[p(o_t \mid s_t)]$





Precisionγをかける．

Softmax関数σで規格化する．
$q(\pi) = \sigma(-\gamma G(\pi))$

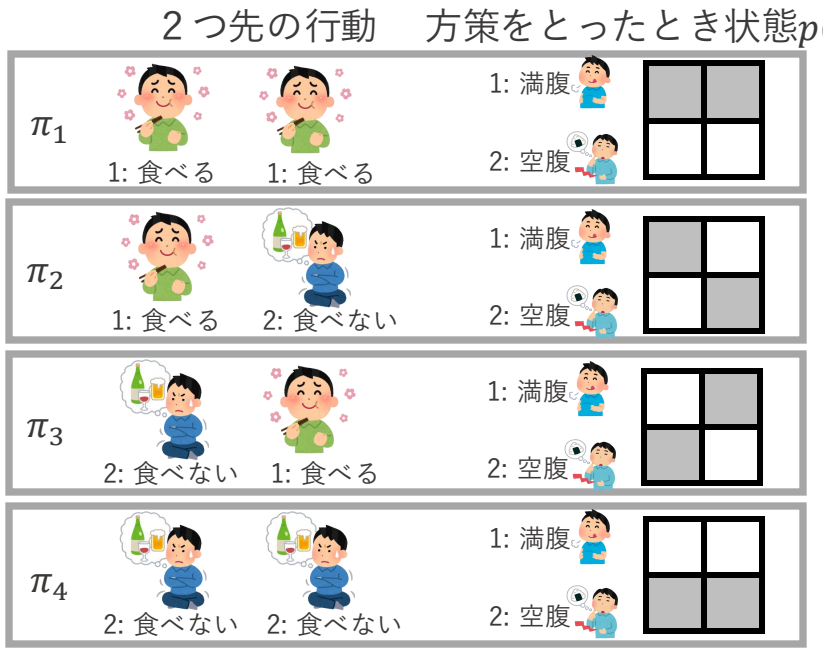


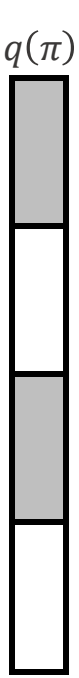
確信を持っていれば，γが大きくなる．よって，探索しなくなる．γはsoftmaxの温度パラメタの逆数だと思えば良い．

期待自由エネルギーを最小にする方策を選ばない

- ここで，期待自由エネルギーを最小にする方策を選ぶこともできる．
- しかし，現在最小にすると思われる方策を取ると，真に最小にする方策を選ぶ機会がなくなる．
- その代わりに，Agentは望む観測を得られやすい方策をとるとする．

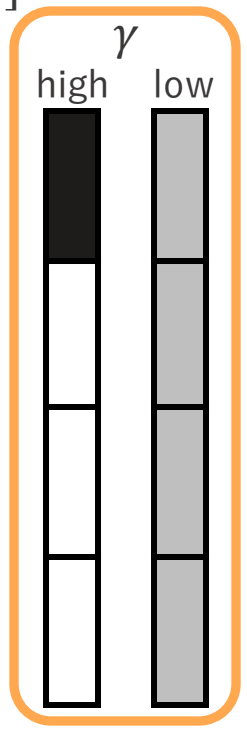
期待自由エネルギー $G = KL\big(q(o_t \mid \pi)||p(o_t)\big) + \sum_s q(s_t \mid \pi) H[p(o_t \mid s_t)]$





Precisionγをかける．

Softmax関数σで規格化する．
$q(\pi) = \sigma(-\gamma G(\pi))$

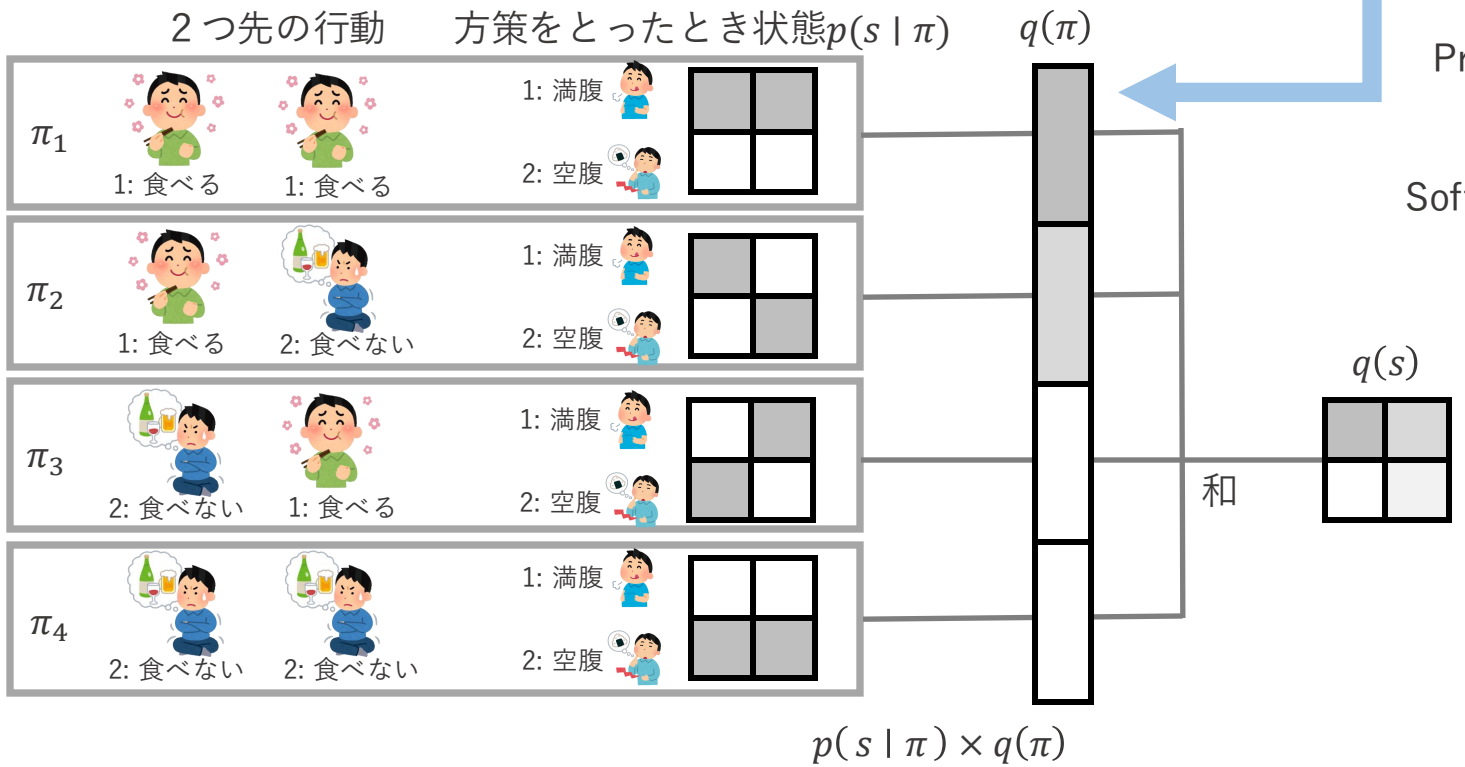


確信を持っていれば，γが大きくなる．よって，探索しなくなる．
γはsoftmaxの温度パラメタの逆数だと思えば良い．

状態$p(s \mid \pi)$と$q(\pi)$の積

- まず，方策で生じる状態$p(s \mid \pi)$と$q(\pi)$の積の和を取る．

期待自由エネルギー $G = KL\big(q(o_t \mid \pi)||p(o_t)\big) + \sum_s q(s_t \mid \pi) H[p(o_t \mid s_t)]$



Precisionγをかける．

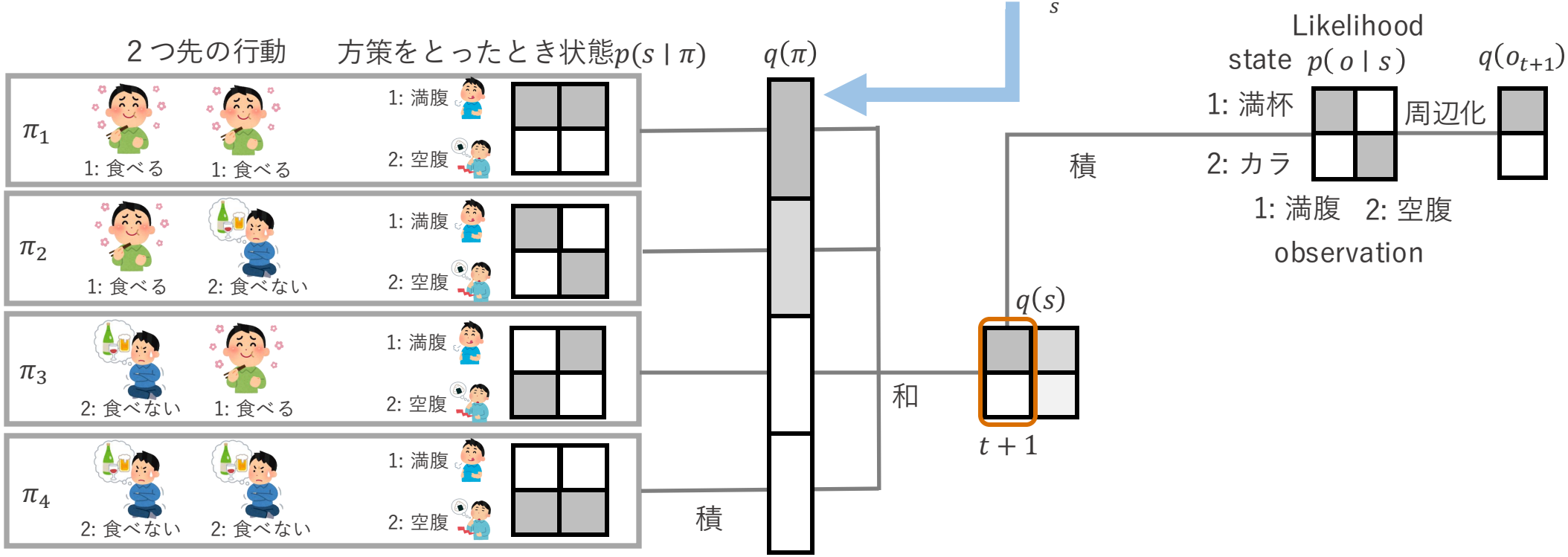
Softmax関数σで規格化する．
$q(\pi) = \sigma(-\gamma G(\pi))$

$q(s|\pi)$の$q(\pi)$の下での期待値，つまり重み付き和をとる．その重みは各方策の確率で定義される。この結果、周辺分布 $q(s)$ が得られる．この分布には方策が暗黙のうちに組み込まれている。

次に生じる観測の予測

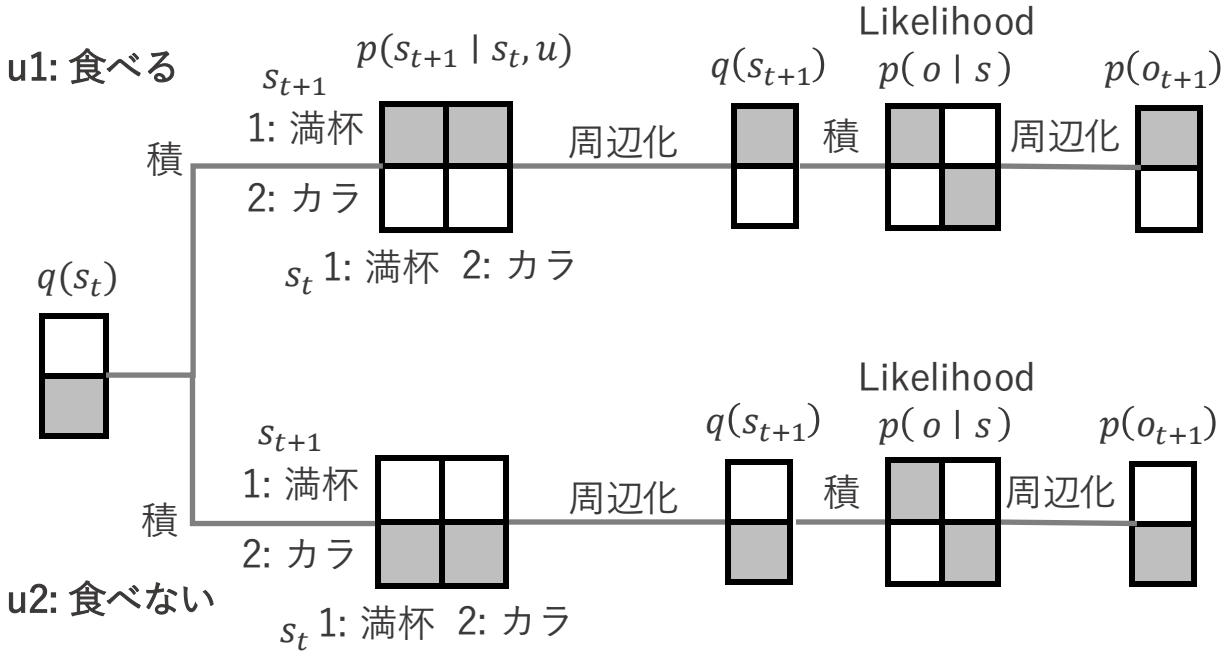
- 次に、期待される観測の確率$q(o_{t+1})$を得るために，次の時間ステップの状態の信念$q(s_{t+1})$に$p(o \mid s)$を掛ける.
- そして，$q(o_{t+1}, s_{t+1})$を周辺化すると$q(o_{t+1})$が求まる.
  - これは期待自由エネルギーから求まった次に生じる観測に対する信念である.

期待自由エネルギー　$G = KL\big(q(o_t \mid \pi)||p(o_t)\big) + \sum_s q(s_t \mid \pi) H[p(o_t \mid s_t)]$



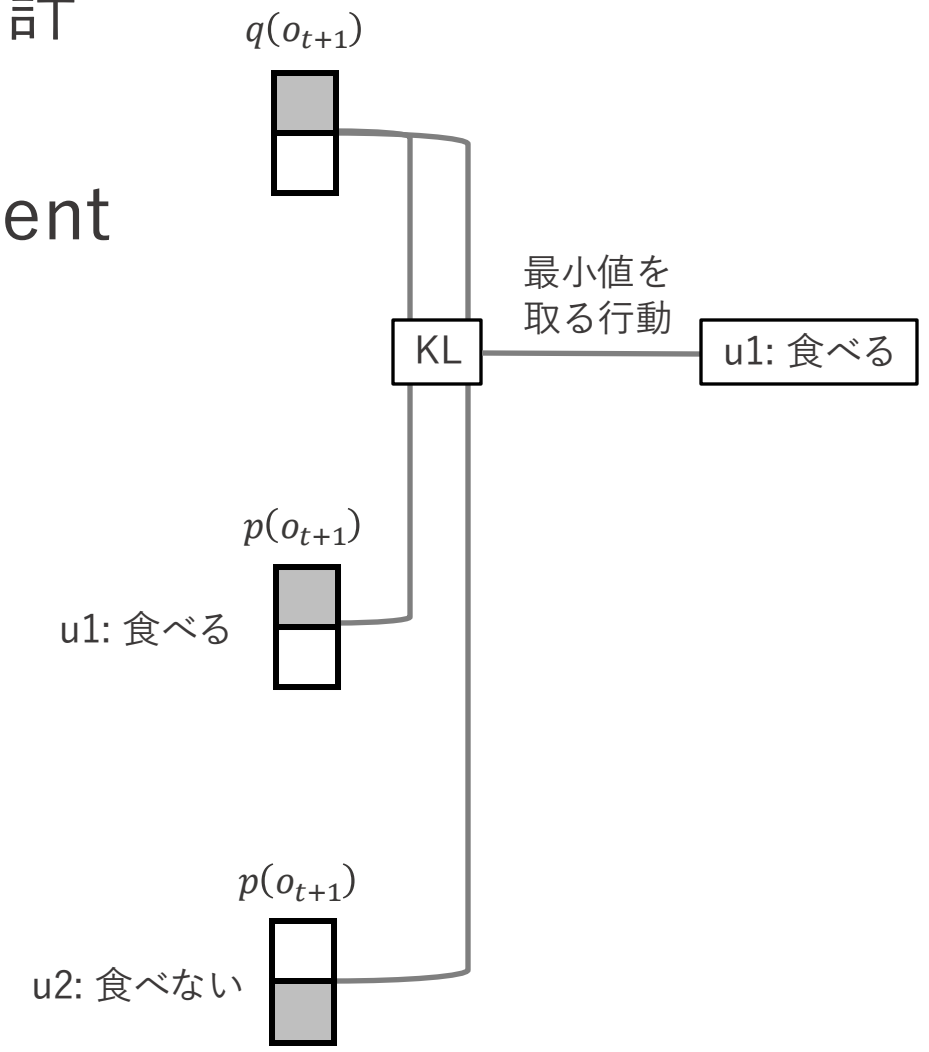
ある行動をとったときに生じる観測

- 現在の状態s_tから行動uをとったときに生じる次の状態s_{t+1}は，$p(s_t \mid s_{t-1}, u)$で決まる．
- まず現在の状態に対する信念$q(s_t)$をとり，行動u1，u2について，次の状態s_{t+1}に対する信念 $q(s_{t+1})$ を求める．
- これを$p(o \mid s)$ にかけて周辺化すると，次の観測の仮説 $p(o_{t+1})$を得る．

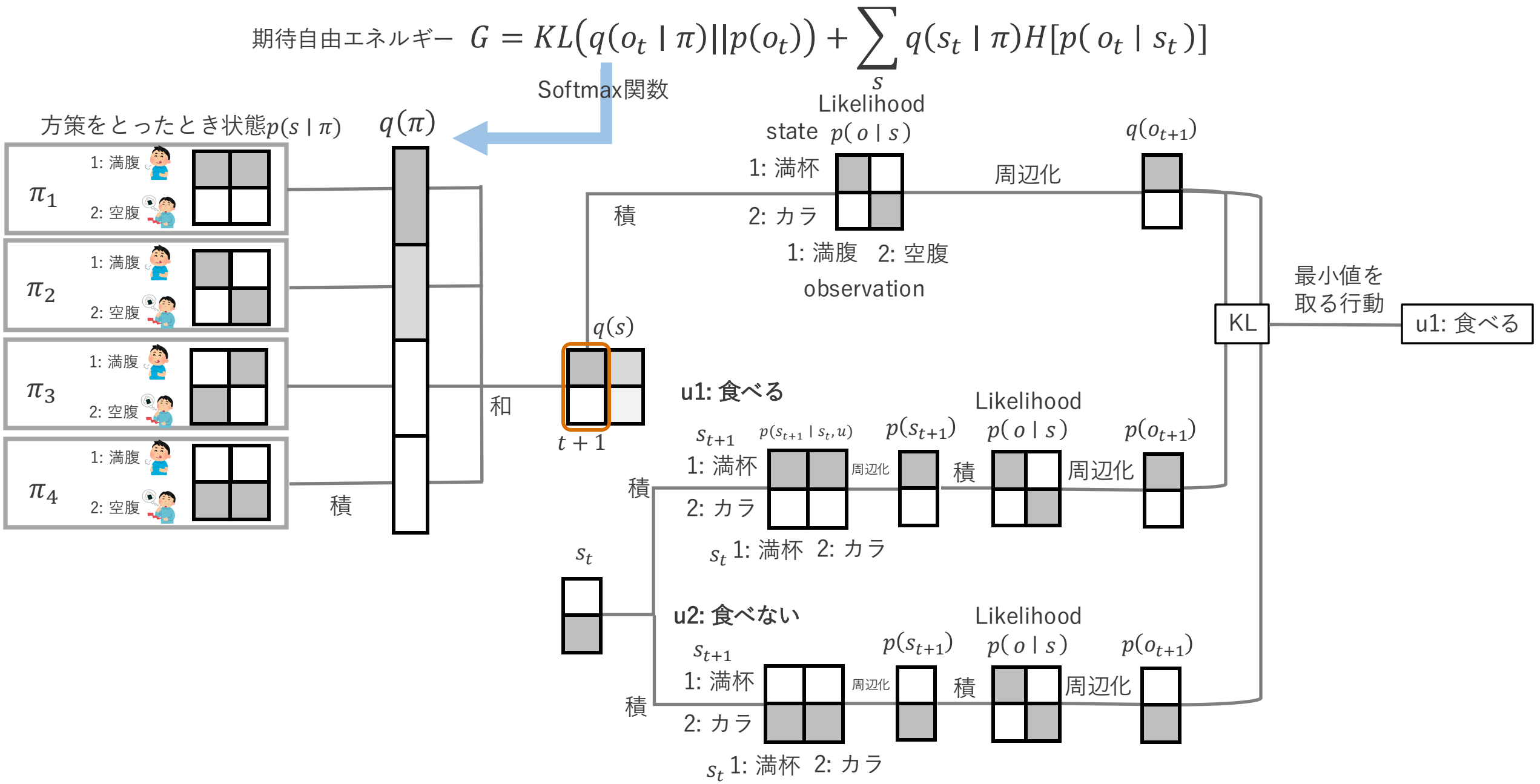


KLダイバージェンスを最小にする行動をとる

- 期待自由エネルギーから求めた$q(o_{t+1})$と，行動から求めた$p(o_{t+1})$のKLダイバージェンスを計算する.
- KLダイバージェンスが最小となる行動をAgentはとる.



まとめの図


期待自由エネルギー $G = KL(q(o_t \mid \pi)||p(o_t)) + \sum_s q(s_t \mid \pi)H[p(o_t \mid s_t)]$
Softmax関数
方策をとったとき状態$p(s \mid \pi)$
$q(\pi)$
π_1
π_2
π_3
π_4
1: 満腹
2: 空腹
積
和
$q(s)$
$t+1$
Likelihood
state $p(o \mid s)$
1: 満杯
2: カラ
1: 満腹 2: 空腹
observation
$q(o_{t+1})$
周辺化
KL
最小値を
取る行動
u1: 食べる
u1: 食べる
s_{t+1}
$p(s_{t+1} \mid s_t, u)$
$p(s_{t+1})$
$p(o \mid s)$
$p(o_{t+1})$
s_t 1: 満杯 2: カラ
s_t
u2: 食べない