Inertial Updating

Adam Dominiak[1] Matthew Kovach[1] Gerelt Tserenjigmid[2]

[1]Virginia Tech
[2]UC Santa Cruz

Indiana University, March 31, 2023

Motivation

- How agents update their beliefs in light of new information is a foundational problem in economics and game theory.
- **Bayesian Updating:** the benchmark model of Bayesian updating has two major issues:
  - **Incomplete:** it is not well-defined for zero-probability events.
    - Ex: dynamic games with incomplete information and refinements of PBE.
  - **Limited:** people systematically deviate from Bayesian updating.
    - Ex: confirmation bias and over- and under-reaction to information.

Motivation

- **Inertial Updating:** A theory of belief updating that overcomes the above two issues:
  - **Complete:** A systematic way of modeling updating under zero-probability events.
  - **Rich:** A unifying framework that nests Bayesian and some well-known non-Bayesian updating rules.

Contributions

- **Inertial updating** is a complete theory of belief updating s.t.
  - DM has a prior μ over S and learns an event $E \subseteq S$,
  - The posterior μ_E is the "closest" element of $\Delta(E)$ from μ:

$$\mu_E = \arg \min_{\pi \in \Delta(E)} d_\mu(\pi).$$

- **Inertial updating** provides a unified framework that nests
  - Bayes' rule and well-known non-Bayesian updating rules such as Grether's (1980) $\alpha - \beta$ rule;
  - Updating rules for zero-probability events such as Myerson's (1986) Conditional Probability System (CPS);
  - Ortoleva's (2012) Hypothesis Testing Model.
- It is characterized by two simple axioms in addition to standard subjective expected utility axioms.

Plan

- Basic setup and model.
- Examples:
  - Bayesian updating;
  - Grether's (1980) $\alpha - \beta$ rule.
- Updating under zero-probability events and Myerson's CPS.
- Ortoleva's (2012) Hypothesis Testing Model.
- Behavioral foundations of Inertial Updating.
- Related Literature.

Basic Setup

- $S = \{s_1, \ldots, s_n\}$ is the set of all states, and $\Delta(S)$ is the set of all probability distributions over S.
- X is the set of all consequences and $\Delta(X)$ is the set of all finite lotteries over X.
- $\mathcal{F} \equiv \{f : S \to \Delta(X)\}$ is the set of all (Anscombe-Aumann) acts.
- Initial SEU preference $\succsim = \succsim_S \Rightarrow$ prior $\mu \in \Delta(S)$.
- Σ is a collection of non-empty subsets of S.
- **Information:** $E \in \Sigma$.
- Conditional SEU preference $\succsim_E \Rightarrow$ posterior $\mu_E \in \Delta(S)$.

Bayesian Updating

- Weather has three outcomes; i.e., $S = \{s, r, h\}$.
- Prior: $\mu = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.
- There will be a storm tomorrow, i.e., $E = \{r, h\} \subset S$.
- Bayesian posterior: $\mu_E = (0, \frac{1}{2}, \frac{1}{2}) \in \Delta(E)$ where

$$\Delta(E) = \{\pi \in \Delta(S) \mid \pi(E) = 1\}.$$

- In other words, DM selects μ_E from $\Delta(E)$.

Bayesian Updating

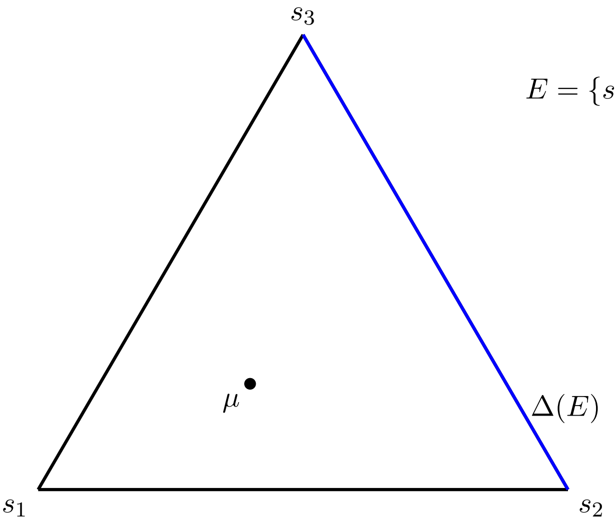
- Weather has three outcomes; i.e., $S = \{s, r, h\}$.
- Prior: $\mu = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.
- There will be a storm tomorrow, i.e., $E = \{r, h\} \subset S$.
- Bayesian posterior: $\mu_E = (0, \frac{1}{2}, \frac{1}{2}) \in \Delta(E)$ where

$$\Delta(E) = \{\pi \in \Delta(S) \mid \pi(E) = 1\}.$$

- In other words, DM selects μ_E from $\Delta(E)$. In fact,

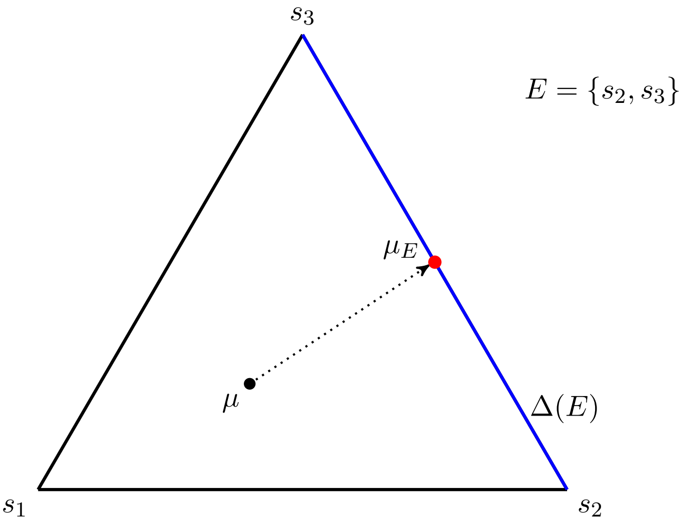
$$\mu_E = \arg \min_{\pi \in \Delta(E)} d_\mu^{KL}(\pi).$$

Bayesian Updating



$E = \{s_2, s_3\}$

Bayesian Updating



Representation

Definition 1

A family of preferences $\{\succsim_E\}_{E\in\Sigma}$ admits an **Inertial Updating** (IU) representation if there exist $\mu \in \Delta(S)$, $u : X \to \mathbb{R}$, and $d : \Delta(S) \to \mathbb{R}$ such that

(i) $\succsim$ is a SEU preference with (μ, u), i.e., for any $f, g \in \mathcal{F}$,

$$f \succsim g \quad \text{iff} \quad \mathbb{E}_\mu u(f) \geq \mathbb{E}_\mu u(g);$$

(ii) for each $E \in \Sigma$, $\succsim_E$ is a SEU preference with (μ_E, u) where

$$\mu_E \equiv \arg\min_{\pi\in\Delta(E)} d(\pi);$$

(iii) μ is the unique minimizer of d.

Examples: Bayesian Updating

Ex 1. (Bayesian Divergence) Let $d^{\sigma}_{\mu}(\pi) = -\sum_{i=1}^{n} \mu_i\, \sigma(\frac{\pi_i}{\mu_i})$ for strictly increasing and strictly concave σ.

- The Kullback-Leibler divergence if $\sigma = \ln$.

Proposition 1

For any E with $\mu(E) > 0$,

$$\mu_E = \arg \min_{\pi \in \Delta(E)} -\sum_{i=1}^{n} \mu_i\, \sigma(\frac{\pi_i}{\mu_i}) = BU(\mu, E).$$

In other words,

$$\mu_E(s) = \frac{\mu(s)}{\mu(E)} \text{ when } s \in E.$$

Examples: Distorted Bayesian Updating

Ex 2. Let $d_\mu(\pi) = -\sum_{i=1}^n h(\mu_i)\,\sigma(\frac{\pi_i}{h(\mu_i)})$. Then

$$\mu_E(s) = \frac{h(\mu(s))}{\sum_{s' \in E} h(\mu(s'))} \text{ when } s \in E.$$

- Bayes' rule when $h(\mu) = \mu$,
- When $h(\mu) = (\mu)^\alpha$, Grether's (1980) $\alpha - \beta$ rule with $\alpha = \beta$.
  - $\alpha < 1$, underreaction to information and base-rate neglect,
  - $\alpha > 1$, overreaction to information and confirmation bias.

Examples: Distorted Bayesian Updating

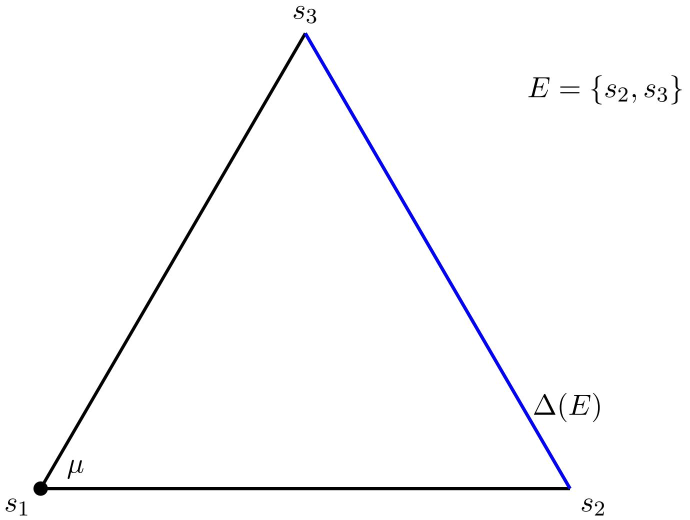
Ex 2. Let $d_\mu(\pi) = -\sum_{i=1}^n h(\mu_i)\,\sigma(\frac{\pi_i}{h(\mu_i)})$. Then

$$\mu_E(s) = \frac{h(\mu(s))}{\sum_{s' \in E} h(\mu(s'))} \text{ when } s \in E.$$

- Bayes' rule when $h(\mu) = \mu$,
- When $h(\mu) = (\mu)^\alpha$, Grether's (1980) $\alpha - \beta$ rule with $\alpha = \beta$.
  - $\alpha < 1$, underreaction to information and base-rate neglect,
  - $\alpha > 1$, overreaction to information and confirmation bias.
- When $h(\mu) = \mu + \delta\,\mathbf{1}\{\mu > \frac{1}{2}\}$, our model generates confirmation bias, similar to Rabin and Schrag (1999).
- We can incorporate history-dependent (or context-dependent) belief updating through h.

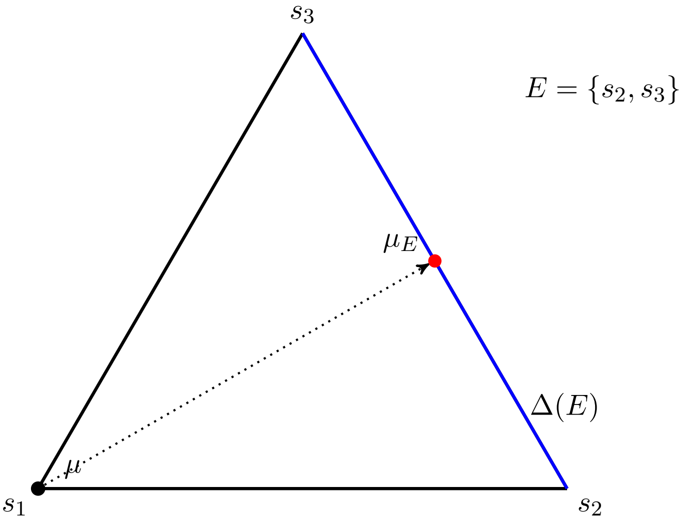
Updating on Zero Probability Events

Suppose $\mu(E) = 0$.



Updating on Zero Probability Events

Suppose $\mu(E) = 0$.



Example: Updating on Zero Probability Events

Ex 3: Let

$$d_\mu(\pi) = \begin{cases} d^\sigma_\mu(\pi) & \text{if } \mu(sp(\pi)) > 0, \\ d^\sigma_{\mu^*}(\pi) + \sigma(1) + |\sigma(0)| & \text{otherwise.} \end{cases}$$

If μ^* has a full-support, then

$$\mu_E = \begin{cases} \text{BU}(\mu, E) & \text{if } \mu(E) > 0, \\ \text{BU}(\mu^*, E) & \text{otherwise.} \end{cases}$$

- The above updating rule was used in Galperti (2019) and is a special case of Myerson's CPS and Ortoleva's hypothesis testing model.

Examples: Non-Bayesian Updating

Ex 4. (Euclidean distance) Let $d_\mu(\pi) = \sum_{i=1}^n (\pi_i - \mu_i)^2$. Then

$$\mu_E(s) = \mu(s) + \frac{1 - \mu(E)}{|E|} \text{ when } s \in E.$$

- Probability is allocated uniformly over states, capturing "$\frac{1}{N}$-heuristic."
- Extends naturally to zero-probability events.

Myerson's CPS and Zero-Probability Events

CPS and Perfect Bayesian Equilibrium (PBE)

- **Incompleteness of Bayes' rule:** In PBE, agents' beliefs are Bayes-consistent with the prior whenever possible. However, there is no restriction when Bayes' rule is not applicable.
- **Refinements:** Refinements of PBE essentially require a **complete theory** of belief updating.
- **Sequential Equilibria of Kreps and Wilson (1982):** Any belief should be a limit of full-support beliefs after applying Bayes' rule accordingly.
- **CPS:** The limit requirement of sequential equilibria is equivalent to the following generalization of Bayes' condition:

$$\mu_E(s) = \mu_F(s)\,\mu_E(F) \text{ for all } s \in F \subseteq E.$$

Definition 2 (CPS)

A conditional probability system is a collection of $\{\mu_E\}_{E\in\Sigma}$ such that

$$\mu_E(s) = \mu_F(s)\,\mu_E(F) \text{ for all } s \in F \subseteq E.$$

- If $\mu_E(F) > 0$, then $\mu_F(s) = \frac{\mu_E(s)}{\mu_E(F)}$ (Bayes' rule).

Proposition 2

Every CPS has an IU representation.

Myerson's CPS

Proposition 3

For any CPS, $\exists \mu^0, \ldots, \mu^K \in \Delta(S)$ *such that* $sp(\mu^0), \ldots, sp(\mu^K)$ *is a partition of* S *and for any* $E \in \Sigma$,

$$\mu_E = BU(\mu^{k^*}, E) \text{ where } k^* = \min\{k \mid \mu^k(E) > 0\}.$$

Moreover, it has an IU representation with the distance function:

$$d_\mu(\pi) = d^\sigma_{\mu^{k^*}}(\pi) + k^* \left(\sigma(1) + |\sigma(0)|\right),$$

where $k^* = \min\{k \mid \mu^k(sp(\pi)) > 0\}$.

Ortoleva's (2012) Hypothesis Testing Model

Ortoleva's (2012) Hypothesis Testing Model

Idea. Apply Bayes' rule if possible. If not, use maximal likelihood.

HTM: DM has a second order prior ρ over $\Delta(\Delta(S))$. For some $\epsilon \in [0,1]$,

$$\mu_E = \begin{cases} \text{BU}(\mu, E) & \text{if } \mu(E) > \epsilon, \\ \text{BU}(\pi_E^\rho, E) & \text{otherwise,} \end{cases}$$

where $\pi_E^\rho = \arg\max_{\pi \in \Delta} \rho(\pi)\pi(E)$.

Ortoleva's Hypothesis Testing Model

Theorem 1

IU and HTM are behaviorally equivalent.

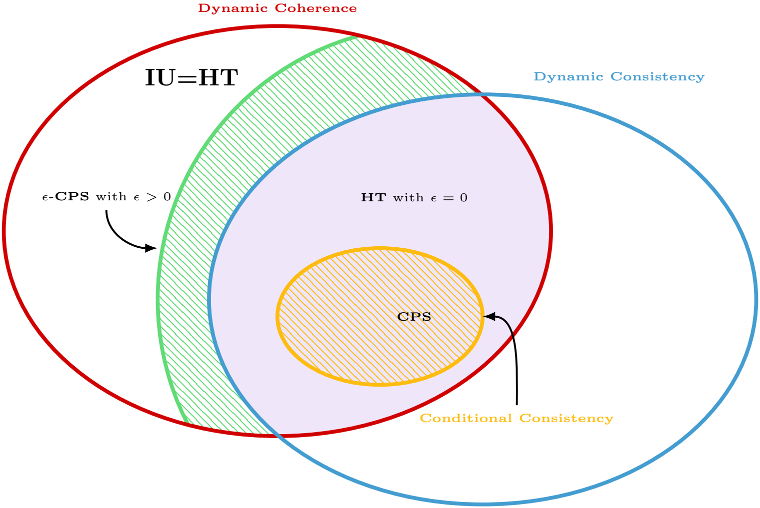
Ortoleva's Hypothesis Testing Model

Theorem 1

IU and HTM are behaviorally equivalent.

Corollary 1

- HTM generalizes Myerson's CPS.
- HTM generalizes Grether's rule.


Dynamic Coherence
IU=HT
Dynamic Consistency
ϵ-**CPS** with $\epsilon > 0$
HT with $\epsilon = 0$
CPS
Conditional Consistency

Behavioral Foundation of IU

Standard Axioms

Axiom 1 (SEU Postulates)

For each $E \in \Sigma$,

(i) **Weak Order:** $\succsim_E$ is complete and transitive;

(ii) **Archimedean:** For any $f, g, h \in \mathcal{F}$, if $f \succ_E g$ and $g \succ_E h$, then there are $\alpha, \beta \in (0,1)$ such that $\alpha f + (1-\alpha)h \succ_E g$ and $g \succ_E \beta f + (1-\beta)h$;

(iii) **Monotonicity:** For any $f, g \in \mathcal{F}$, if $f(s) \succsim_E g(s)$ for each $s \in S$, then $f \succsim_E g$;

(iv) **Nontriviality:** There are $f, g \in \mathcal{F}$ such that $f \succ_E g$;

(v) **Independence:** For any $f, g, h \in \mathcal{F}$ and $\alpha \in (0,1]$, $f \succsim_E g$ if and only if $\alpha f + (1-\alpha)h \succsim_E \alpha g + (1-\alpha)h$.

(vi) **Invariant Risk Preference:** For any lotteries $p, q \in \Delta(X)$, $p \succsim_E q$ if and only if $p \succsim q$.

Standard Axioms

Axiom 2 (Consequentialism)

For any E and $f, g \in \mathcal{F}$,

$$f(s) = g(s) \text{ for all } s \in E \Rightarrow f \sim_E g.$$

Standard Axioms

Axiom 2 (Consequentialism)

For any E and $f, g \in \mathcal{F}$,

$$f(s) = g(s) \text{ for all } s \in E \Rightarrow f \sim_E g.$$

Lemma 1. SEU postulates imply that there are μ, u, and $\{\mu_E\}_{E \in \Sigma}$ such that

- $\succsim$ has a SEU representation with (μ, u),
- $\succsim_E$ has a SEU representation with (μ_E, u).

Standard Axioms

Axiom 2 (Consequentialism)

For any E and $f, g \in \mathcal{F}$,

$$f(s) = g(s) \text{ for all } s \in E \Rightarrow f \sim_E g.$$

Lemma 1. SEU postulates imply that there are μ, u, and $\{\mu_E\}_{E \in \Sigma}$ such that

- $\succsim$ has a SEU representation with (μ, u),
- $\succsim_E$ has a SEU representation with (μ_E, u).

Lemma 2. SEU postulates and Consequentialism imply that $\mu_E \in \Delta(E)$.

Substantive Axiom

Revealed Preference: An event A is **revealed implied** by an event B if $S \backslash A$ is $\succsim_B$-null; i.e.,

$$(p, A; q, S \backslash A) \sim_B p \text{ for any } p, q \in \Delta(X).$$

In words, every state that the DM believes is possible after learning B is in A.

Axiom 3 (Dynamic Coherence)

For any $A_1, \ldots, A_n \subseteq S$, if $S \setminus A_i$ is $\succsim_{A_{i+1}}$-null for each $i \leq n-1$ and $S \setminus A_n$ is $\succsim_{A_1}$-null, then $\succsim_{A_1} = \succsim_{A_n}$.

Behavioral Foundation of IU

Theorem 2

The following are equivalent.

(i) *A family of preferences* $\{\succsim_E\}_{E \in \Sigma}$ *satisfies* **SEU Postulates**, **Consequentialism**, *and* **Dynamic Coherence**.

(ii) *It admits an* **IU** *representation.*

(iii) *It admits an* **IU** *representation with respect to a continuous, strictly convex distance function.*

Proof Sketch

- **Step 1.** SEU axioms imply that $\succsim$ is a SEU preference with some (μ, u) and $\succsim_E$ is a SEU preference with some (μ_E, u).
- **Step 2.** Consequentialism implies that $\mu_E \in \Delta(E)$.

Proof Sketch

- **Step 1.** SEU axioms imply that $\succsim$ is a SEU preference with some (μ, u) and $\succsim_E$ is a SEU preference with some (μ_E, u).
- **Step 2.** Consequentialism implies that $\mu_E \in \Delta(E)$.
- Note that $S \setminus A$ is $\succsim_B$-null means that $\mu_B(S \setminus A) = 0$; equivalently, $\mu_B(A) = 1$. In other words, A **is revealed implied by** B iff $\mu_B \in \Delta(A)$.
- In other words, μ_B is chosen from $\Delta(A)$ in the presence of μ_B. Hence, Dynamic Coherence is equivalent to SARP on this revealed preference.
- **Step 3.** By an extension of Afriat's (1967) theorem for general budget sets due to Matzkin (1991), $\exists v$ s.t

$$\mu_E = \arg \max_{\pi \in \Delta(E)} v(\pi).$$

Related Literature

IU representation:

- Perea (2009) axiomatizes Euclidean distance based updating rules.
  - Always non-Bayesian.
- Basu (2019) characterizes lexicographic (AGM-consistent) updating rules and he shows that a special case of these rules also has an IU representation.
  - Bayesian whenever possible.

Non-Bayesian updating rules: Epstein (2006), Epstein et al. (2008), and Kovach (2020).

- They deviate from Consequentialism.

Conclusion

- We propose and characterize a **complete theory for belief updating** that overcomes two major issues of Bayesian updating.
- In IU, the DM selects the "closest" belief to her prior given information.
- IU nests Bayesian updating and some well-known non-Bayesian updating rules.
    - Grether's (1980) $\alpha - \beta$ rule.
    - Myerson's (1986) CPS.
    - Ortoleva's (2012) Hypothesis Testing Model.

Additional results

- Apply our model to the signal structure.
- Axiomatic Characterization of CPS.
- Apply our model to the Bayesian persuasion game and illustrate the effect of belief distortions on the optimal signal structures.
- $\epsilon-$CPS is a non-Bayesian extension of CPS that is still a special case of IU.
- Relaxing Consequentialism:

$$\mu_E = \delta\,\mu + (1-\delta)\ \arg\min_{\pi\in\Delta(E)} d_\mu(\pi).$$

Thank you!!

gtserenj@ucsc.edu

Signal Structure – Bayesian Updating

- Let Ω be the payoff relevant state space and M be the set of all signals. Then let $S = \Omega \times M$.
- Let $P(\omega)$ be the (unconditional) probability that the payoff relevant state ω occurs.
- Let $P(m|\omega)$ be the (conditional) probability that the DM receives the signal m when the state is ω.
- Note that receiving signal m is equivalent to learning the event $\{(\omega, m)\}_{\omega \in \Omega}$ in S.
- Let μ be the prior on S: $\mu_{\omega m} = P(m|\omega)\, P(\omega)$.
- The Bayesian divergence generates Bayesian updating in the signal structure framework:

$$P(\omega|m) = \frac{\mu_{\omega m}}{\sum_{\omega' \in \Omega} \mu_{\omega' m}} = \frac{P(m|\omega)\, P(\omega)}{\sum_{\omega' \in \Omega} P(m|\omega')\, P(\omega')}.$$

Signal Structure – Non-Bayesian Updating

- Consider

$$d_\mu(\pi) = \sum_{(\omega,m)\in\Omega\times M} \left(\sum_{m'\in M} \mu_{\omega m'}\right)^{\beta-\alpha} \mu_{\omega m}^{\alpha}\, \sigma\left(\frac{\pi_{\omega m}}{\mu_{\omega m}^{\alpha}}\right).$$

- This distance generates Grether's (1980) $\alpha - \beta$ rule:

$$P(\omega|m) = \frac{\left(\sum_{m'\in M} \mu_{\omega m'}\right)^{\beta-\alpha} \mu_{\omega m}^{\alpha}}{\sum_{\omega'\in\Omega} \left(\sum_{m'\in M} \mu_{\omega' m'}\right)^{\beta-\alpha} \mu_{\omega' m}^{\alpha}}$$

$$= \frac{(P(m|\omega))^{\alpha}\,(P(\omega))^{\beta}}{\sum_{\omega'\in\Omega}(P(m|\omega'))^{\alpha}\,(P(\omega'))^{\beta}}.$$