Dynamically Aggregating Diverse Information

Annie Liang[1] Xiaosheng Mu[2] Vasilis Syrgkanis[3]

[1]University of Pennsylvania

[2]Columbia University [3]Microsoft Research

Introduction

- in many learning problems, don't have access to information about exactly what you want to know
- instead aggregate related information
- e.g. suppose a hotel chain wants to forecast demand for a new location in Puerto Rico
- can't directly learn about this, but can learn about different components:
  - website traffic to the Puerto Rico tourism bureau provides estimate of tourism travel
  - Google search data for local conference venues provides estimate of business travel
- improve forecasting by aggregating this data
- how to acquire data over time, given limited resources?

Our Model

- decision-maker (DM) seeks to learn Gaussian state

Our Model

- decision-maker (DM) seeks to learn Gaussian state (volume of travel to hypothetical new location in Puerto Rico)

Our Model

- decision-maker (DM) seeks to learn Gaussian state (volume of travel to hypothetical new location in Puerto Rico)

- state is a linear combination of unknown attributes

Our Model

- decision-maker (DM) seeks to learn Gaussian state (volume of travel to hypothetical new location in Puerto Rico)
- state is a linear combination of unknown attributes (vacation travel, business travel, etc.)

Our Model

- decision-maker (DM) seeks to learn Gaussian state (volume of travel to hypothetical new location in Puerto Rico)
- state is a linear combination of unknown attributes (vacation travel, business travel, etc.)
- attributes are potentially correlated

Our Model

- decision-maker (DM) seeks to learn Gaussian state (volume of travel to hypothetical new location in Puerto Rico)
- state is a linear combination of unknown attributes (vacation travel, business travel, etc.)
- attributes are potentially correlated (vacation travel from US, vacation travel from Mexico)

Our Model

- decision-maker (DM) seeks to learn Gaussian state (volume of travel to hypothetical new location in Puerto Rico)
- state is a linear combination of unknown attributes (vacation travel, business travel, etc.)
- attributes are potentially correlated (vacation travel from US, vacation travel from Mexico)
- DM has access to a diffusion process about each attribute, allocates attention across them

Our Model

- decision-maker (DM) seeks to learn Gaussian state (volume of travel to hypothetical new location in Puerto Rico)
- state is a linear combination of unknown attributes (vacation travel, business travel, etc.)
- attributes are potentially correlated (vacation travel from US, vacation travel from Mexico)
- DM has access to a diffusion process about each attribute, allocates attention across them (e.g. employee hours)

Our Model

- decision-maker (DM) seeks to learn Gaussian state (volume of travel to hypothetical new location in Puerto Rico)
- state is a linear combination of unknown attributes (vacation travel, business travel, etc.)
- attributes are potentially correlated (vacation travel from US, vacation travel from Mexico)
- DM has access to a diffusion process about each attribute, allocates attention across them (e.g. employee hours)
- at chosen time, stops information acquisition and takes action

Our Model

- decision-maker (DM) seeks to learn Gaussian state (volume of travel to hypothetical new location in Puerto Rico)
- state is a linear combination of unknown attributes (vacation travel, business travel, etc.)
- attributes are potentially correlated (vacation travel from US, vacation travel from Mexico)
- DM has access to a diffusion process about each attribute, allocates attention across them (e.g. employee hours)
- at chosen time, stops information acquisition and takes action (whether or not to open new location in Puerto Rico)

Preview of Main Result

under assumption on prior belief (over attributes), optimal information acquisition is “simple”

- DM initially focuses all attention on one attribute
- progressively adds in new attributes
- constant attention allocation during each stage
- strategy is history-independent

Preview of Main Result

under assumption on prior belief (over attributes), optimal information acquisition is "simple"

- DM initially focuses all attention on one attribute
- progressively adds in new attributes
- constant attention allocation during each stage
- strategy is history-independent

and "robust":

- optimal across large class of payoff/cost specifications

Preview of Main Result

under assumption on prior belief (over attributes), optimal information acquisition is "simple"

- DM initially focuses all attention on one attribute
- progressively adds in new attributes
- constant attention allocation during each stage
- strategy is history-independent

and "robust":

- optimal across large class of payoff/cost specifications

applications to: binary choice, competing information providers

Plan for Talk

1. Model
2. Two Attributes
3. Many Attributes
4. Application: Binary Choice
5. Application: Competing Information Providers

Informational Environment

unknown attributes	$\theta = (\theta_1, \ldots, \theta_K) \sim \mathcal{N}(\mu, \Sigma)$
payoff-relevant state	$\omega = \sum_{i=1}^K \alpha_i \theta_i$ with each $\alpha_i > 0$
data sources	diffusion process X_i for each θ_i

Attention Allocation

- continuous time $t \in \mathbb{R}_+$
- allocate unit of attention across attributes at each time t

  $(\beta_1^t, \ldots, \beta_K^t)$ where $\sum_{i=1}^K \beta_i^t = 1$
- attention choices influence the diffusion processes via

  $$dX_i^t = \beta_i^t \cdot \theta_i \cdot dt + \sqrt{\beta_i^t} \cdot dB_i^t$$

  where B_i are independent standard Brownian motions.
- DM observes complete paths of each process: at each time t the history is $\left\{X_i^{\leq t}\right\}_{i=1}^K$

Decision Problem

DM chooses

- information acquisition strategy S: map from histories into an attention vector
- stopping rule τ: map from history into decision of whether to stop sampling

Criterion:

$$\max_{S,\tau} \mathbb{E}\left[\max_{a} \mathbb{E}[u(a,\omega) \mid \mathcal{F}_\tau] - c(\tau)\right]$$

for some arbitrary positive increasing cost function c.

Comments on Problem

results will characterize optimal information acquisition only

- in general, S and τ would have to be determined jointly
- we show that they can be separated under a condition on the prior belief

Comments on Problem

results will characterize optimal information acquisition only

- in general, S and τ would have to be determined jointly
- we show that they can be separated under a condition on the prior belief

this is not a multi-armed bandit problem

- in MAB, actions play the dual role of influencing the evolution of beliefs and determining flow payoffs
- here they are separated
- so information acquisition decisions are driven by learning concerns exclusively

Related Literature

- **Dynamic Learning from Fixed Set of Signals:**
  Moscarini-Smith ('01), Fudenberg et al. ('18), Che-Mierendorff ('19), Mayskaya ('19)

- **Rational Inattention and Flexible Information Acquisition:**
  Steiner et al. ('09), Hébert-Woodford ('18), Zhong ('18)

- **Statistics:**
  multi-armed bandits; optimal experiment design; comparison of experiments.

Related Literature

- Dynamic Learning from Fixed Set of Signals:
  Moscarini-Smith ('01), Fudenberg et al. ('18), Che-Mierendorff ('19), Mayskaya ('19)

  $\longrightarrow$ we allow many signals with flexible correlation

- Rational Inattention and Flexible Information Acquisition:
  Steiner et al. ('09), Hébert-Woodford ('18), Zhong ('18)

- Statistics:
  multi-armed bandits; optimal experiment design; comparison of experiments.

Related Literature

- Dynamic Learning from Fixed Set of Signals:
  Moscarini-Smith ('01), Fudenberg et al. ('18), Che-Mierendorff ('19), Mayskaya ('19)
  $\longrightarrow$ we allow many signals with flexible correlation
- Rational Inattention and Flexible Information Acquisition:
  Steiner et al. ('09), Hébert-Woodford ('18), Zhong ('18)
  $\longrightarrow$ our signals and information cost are "fixed"
- Statistics:
  multi-armed bandits; optimal experiment design; comparison of experiments.

Related Literature

- Dynamic Learning from Fixed Set of Signals:
  Moscarini-Smith ('01), Fudenberg et al. ('18), Che-Mierendorff ('19), Mayskaya ('19)
  $\longrightarrow$ we allow many signals with flexible correlation

- Rational Inattention and Flexible Information Acquisition:
  Steiner et al. ('09), Hébert-Woodford ('18), Zhong ('18)
  $\longrightarrow$ our signals and information cost are "fixed"

- Statistics:
  multi-armed bandits; optimal experiment design; comparison of experiments.
  $\longrightarrow$ our model closest to recent work on "best-arm identification";
  solves "identification" between two correlated Gaussian arms

Two Sources ($K = 2$)

Two Sources

- two unknown attributes

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

- access to two Brownian motions

- agent seeks to learn $\omega = \alpha_1\theta_1 + \alpha_2\theta_2$, where each $\alpha_i > 0$.

Key Condition on Prior Beliefs

define $y_1 = \alpha_1 \Sigma_{11} + \alpha_2 \Sigma_{12}$ and $y_2 = \alpha_1 \Sigma_{21} + \alpha_2 \Sigma_{22}$.

Assumption

The prior covariance matrix satisfies $y_1 + y_2 \geq 0$.

Key Condition on Prior Beliefs

define $y_1 = \alpha_1 \Sigma_{11} + \alpha_2 \Sigma_{12}$ and $y_2 = \alpha_1 \Sigma_{21} + \alpha_2 \Sigma_{22}$.

Assumption

The prior covariance matrix satisfies $y_1 + y_2 \geq 0$.

loosely, this requires the two attributes to be not too negatively correlated

- always satisfied if $\alpha_1 = \alpha_2$
  $\longrightarrow$ agent wants to learn $\omega = \theta_1 + \theta_2$
- or $\Sigma_{12} = \Sigma_{21} \geq 0$
  $\longrightarrow$ attributes are positively correlated
- or $\Sigma_{11} = \Sigma_{22}$
  $\longrightarrow$ same initial uncertainty about the two attributes

Optimal Attention Allocation Strategy

Theorem

Wlog let $y_1 \geq y_2$. Define

$$t_1 = \frac{y_1 - y_2}{\alpha_2 \det(\Sigma)}.$$

Optimal Attention Allocation Strategy

Theorem

Wlog let $y_1 \geq y_2$. Define

$$t_1 = \frac{y_1 - y_2}{\alpha_2 \det(\Sigma)}.$$

Under the previous assumption, the optimal attention strategy has two stages:

Optimal Attention Allocation Strategy

Theorem

Wlog let $y_1 \geq y_2$. Define

$$t_1 = \frac{y_1 - y_2}{\alpha_2 \det(\Sigma)}.$$

Under the previous assumption, the optimal attention strategy has two stages:

1. *At times $t \leq t_1$, DM optimally attends only to attribute 1.*

Optimal Attention Allocation Strategy

Theorem

Wlog let $y_1 \geq y_2$. Define

$$t_1 = \frac{y_1 - y_2}{\alpha_2 \det(\Sigma)}.$$

Under the previous assumption, the optimal attention strategy has two stages:

1. *At times $t \leq t_1$, DM optimally attends only to attribute 1.*
2. *At times $t > t_1$, DM allocates attention in the constant fraction*

$$(\beta_1^t, \beta_2^t) = \left(\frac{\alpha_1}{\alpha_1 + \alpha_2}, \frac{\alpha_2}{\alpha_1 + \alpha_2}\right).$$

Example 1: Independent Attributes

- unknown attributes

  $$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

  want to learn $\theta_1 + \theta_2$

- then optimally:
  - phase 1: put all attention on learning about θ_1
  - at time $t = 5/6$, posterior covariance matrix is $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
  - after, split attention equally

Example 2: Correlated Attributes

- unknown attributes

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} 6 & 2 \\ 2 & 1 \end{pmatrix} \right)$$

  want to learn $\theta_1 + \theta_2$

- then optimally:
  - phase 1: put all attention on learning about θ_1
  - at $t = 5/2$, posterior covariance is $\begin{pmatrix} 3/8 & 1/8 \\ 1/8 & 3/8 \end{pmatrix}$
  - after, split attention equally

Example 2: Unequal Payoff Weights

- unknown attributes

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \right)$$

  want to learn $\theta_1 + 2\theta_2$

- then optimally:
  - phase 1: put all attention on learning about θ_1
  - at $t = 3/2$, posterior covariance is $\begin{pmatrix} 3/5 & 1/5 \\ 1/5 & 2/5 \end{pmatrix}$
  - after, split attention in the vector $(1/3, 2/3)$

Interpretation of Strategy

Stage 1

Put all attention on learning about attribute 1, where by assumption: $y_1 = \alpha_1 \Sigma_{11} + \alpha_2 \Sigma_{12} \geq \alpha_1 \Sigma_{21} + \alpha_2 \Sigma_{22} = y_2$.

suppose equal payoff weights ($\alpha_1 = \alpha_2$) or independent attributes ($\Sigma_{12} = \Sigma_{21} = 0$)

- above expression reduces to $\Sigma_{11} \geq \Sigma_{22}$
- direct comparison of which attribute the DM is initially more uncertain about
- focus on the attribute with greater initial uncertainty

Interpretation of Strategy

Stage 1

Put all attention on learning about attribute 1, where by assumption: $\alpha_1 \Sigma_{11} + \alpha_2 \Sigma_{12} \geq \alpha_1 \Sigma_{21} + \alpha_2 \Sigma_{22}$.

with unequal payoff weights, want to "re-weight" uncertainty in proportion to those weights:

- higher $\alpha_1 \Rightarrow$ greater value to learning about attribute 1

with correlation:

- learning about attribute 1 has value also in teaching about attribute 2 (and vice versa)

Interpretation of Strategy

- eventually DM has equal (payoff-reweighted) uncertainty about the two attributes

Interpretation of Strategy

- eventually DM has equal (payoff-reweighted) uncertainty about the two attributes

Stage 2

Devote attention in constant fraction $\left(\frac{\alpha_1}{\alpha_1+\alpha_2}, \frac{\alpha_2}{\alpha_1+\alpha_2}\right)$.

Interpretation of Strategy

- eventually DM has equal (payoff-reweighted) uncertainty about the two attributes

Stage 2

Devote attention in constant fraction $\left(\frac{\alpha_1}{\alpha_1+\alpha_2}, \frac{\alpha_2}{\alpha_1+\alpha_2}\right)$.

- these weights produce an unbiased signal about ω:

$$\frac{\alpha_1}{\alpha_1+\alpha_2} \cdot \theta_1 + \frac{\alpha_2}{\alpha_1+\alpha_2} \cdot \theta_2 = \frac{1}{\alpha_1+\alpha_2} \cdot \omega$$

- efficient aggregation of information in "prior-free" sense
- acquisition of signals in this mixture maintains equivalence of marginal values

Conceptual Takeaways

optimal information acquisition is “simple”:

- attention allocations do not depend on the history of signal realizations
- DM can map out and implement a deterministic plan for information acquisition from time 0
- note: expect stopping time and optimal action a to depend on signal realizations

and “robust”:

- strategy does not depend on payoff function $u(a, \omega)$
- note: important that the payoff-relevant state does not change

Practical Takeaways

closed-form expressions for optimal information acquisition strategy in this environment

can use this to:

- characterize exact information acquisition strategy
- study various comparative statics (example later)
- simplify larger problems where information acquisition is not the direct object of interest (example later)

General K

Generalized Condition on Prior

Assumption

The prior covariance matrix satisfies

$$|\Sigma_{ij}| \leq \frac{1}{2K-3} \cdot \Sigma_{ii}, \quad \forall i \neq j.$$

Generalized Condition on Prior

Assumption

The prior covariance matrix satisfies

$$|\Sigma_{ij}| \leq \frac{1}{2K-3} \cdot \Sigma_{ii}, \quad \forall i \neq j.$$

- limits size of covariances (relative to variances)
- for case of $K = 2$, reduces to $|\Sigma_{ij}| \leq \Sigma_{ii}$ (covariances smaller than variances), which implies previous condition for $K = 2$
- condition becomes more stringent for larger K

Optimal Information Acquisition Strategy

Theorem

Under the preceding assumption, there are (up to) K stages of information acquisition, identified with the increasing times

$$0 = t_0 \leq t_1 \leq \cdots \leq t_{K-1} < t_K = +\infty$$

and nested sets

$$\emptyset = B_0 \subsetneq B_1 \subset \cdots B_{K-1} \subsetneq B_K = \{1, \ldots, K\}.$$

Optimal Information Acquisition Strategy

Theorem

Under the preceding assumption, there are (up to) K stages of information acquisition, identified with the increasing times

$$0 = t_0 \leq t_1 \leq \cdots \leq t_{K-1} < t_K = +\infty$$

and nested sets

$$\emptyset = B_0 \subsetneq B_1 \subset \cdots B_{K-1} \subsetneq B_K = \{1, \ldots, K\}.$$

At each stage $[t_{k-1}, t_k)$:

- *the optimal attention level is constant*
- *and supported on the sources in B_k.*

Optimal Information Acquisition Strategy

Theorem

Under the preceding assumption, there are (up to) K stages of information acquisition, identified with the increasing times

$$0 = t_0 \leq t_1 \leq \cdots \leq t_{K-1} < t_K = +\infty$$

and nested sets

$$\emptyset = B_0 \subsetneq B_1 \subset \cdots B_{K-1} \subsetneq B_K = \{1, \ldots, K\}.$$

At each stage $[t_{k-1}, t_k)$*:*

- *the optimal attention level is constant*
- *and supported on the sources in* B_k.

At the final stage, attention is proportional to the weight vector α.

Example

- unknown attributes

$$\begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & -1 \\ 0 & -1 & 3 \end{pmatrix} \right)$$

  want to learn $\omega = \theta_1 + \theta_2 + \theta_3$

- then optimally:
  - phase 1: put all attention on learning about θ_1

Example

- unknown attributes

$$\begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & -1 \\ 0 & -1 & 3 \end{pmatrix} \right)$$

  want to learn $\omega = \theta_1 + \theta_2 + \theta_3$

- then optimally:
  - phase 1: put all attention on learning about θ_1
    - at $t = 1/12$, marginal values of θ_1 and θ_2 have equalized

Example

- unknown attributes

$$\begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & -1 \\ 0 & -1 & 3 \end{pmatrix} \right)$$

  want to learn $\omega = \theta_1 + \theta_2 + \theta_3$

- then optimally:
  - phase 1: put all attention on learning about θ_1
    - at $t = 1/12$, marginal values of θ_1 and θ_2 have equalized
  - phase 2: divide attention between θ_1 and θ_2 in constant mixture $(4/7, 3/7)$
    - at $t = 13/44$, all three marginal values are the same

Example

- unknown attributes

$$\begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & -1 \\ 0 & -1 & 3 \end{pmatrix} \right)$$

  want to learn $\omega = \theta_1 + \theta_2 + \theta_3$

- then optimally:
  - phase 1: put all attention on learning about θ_1
    - at $t = 1/12$, marginal values of θ_1 and θ_2 have equalized
  - phase 2: divide attention between θ_1 and θ_2 in constant mixture $(4/7, 3/7)$
    - at $t = 13/44$, all three marginal values are the same
  - phase 3: split attention equally across sources

Some Properties of the Optimal Strategy

- step-like structure:
  - once DM starts acquiring information from a source, always acquires information from that source
  - progressively adds in new sources
- at each stage, information acquisition is constant
- the times t_k and sets B_k are "history-independent": can be mapped out from $t = 0$
- strategy holds for all payoff functions $u(a, \omega)$

Proof Sketch 1/4: Preliminaries

- at every time t, past attention levels integrate to a cumulated attention vector $q(t) = (q_1(t), \ldots, q_K(t))$
- describes how much attention has been paid to each attribute

Proof Sketch 1/4: Preliminaries

- at every time t, past attention levels integrate to a cumulated attention vector $q(t) = (q_1(t), \ldots, q_K(t))$
- describes how much attention has been paid to each attribute
- let $V(q(t))$ be the posterior variance of ω at time t

Proof Sketch 1/4: Preliminaries

- at every time t, past attention levels integrate to a cumulated attention vector $q(t) = (q_1(t), \ldots, q_K(t))$
- describes how much attention has been paid to each attribute
- let $V(q(t))$ be the posterior variance of ω at time t
- warm-up: suppose there is a fixed stopping time T

Proof Sketch 1/4: Preliminaries

- at every time t, past attention levels integrate to a cumulated attention vector $q(t) = (q_1(t), \ldots, q_K(t))$
- describes how much attention has been paid to each attribute
- let $V(q(t))$ be the posterior variance of ω at time t
- warm-up: suppose there is a fixed stopping time T
- $q(T)$ should minimize $V(q)$ among all vectors q that allocate T units of attention (Hansen-Torgensen)

Proof Sketch 1/4: Preliminaries

- at every time t, past attention levels integrate to a cumulated attention vector $q(t) = (q_1(t), \ldots, q_K(t))$
- describes how much attention has been paid to each attribute
- let $V(q(t))$ be the posterior variance of ω at time t
- warm-up: suppose there is a fixed stopping time T
- $q(T)$ should minimize $V(q)$ among all vectors q that allocate T units of attention (Hansen-Torgensen)
- (note: "order" doesn't matter, just need to integrate to best cumulated attention vector at time T)

Proof Sketch 2/4: Uniform Optimality

Definition

For each time t, define the t-optimal attention vector

$$n(t) := \operatorname*{argmin}_{q\,:\,\|q\|_1 = t} V(q)$$

- suppose it were possible to achieve $n(t)$ at every t
  $\implies$ minimize posterior variance at every time t
- call such a strategy is uniformly optimal
- if a uniformly optimal strategy exists, it is optimal for all payoff criteria (Greenshtein)
- key question is whether a uniformly optimal strategy exists.

Proof Sketch 3/4: Monotonicity of $n(t)$

- sufficient and necessary condition: $n(t)$ weakly increases in t in all coordinates.
- in this case, optimal attention levels β^t are simply the time derivatives of $n(t)$
- when might this fail? example
  - strong complementarity/substitutability across signals
  - locally best reductions in variance need not be best given opportunity to acquire information on a larger time interval
- work with the Hessian of the posterior variance function V
- condition on prior limits extent to which learning about attribute i affects value to attribute j (size of cross-partial)

Proof Sketch (4/4): Step Structure

- at each stage k, agent optimally divides attention among the k attributes in B_k
- specific mixture of information maintains equivalence of marginal values of those k attributes
- reduces the marginal value of each of the k attributes
- eventually, some new attribute will have the same marginal value as the first k attributes
- at this point the agent expands his observation set to include that new attribute
- repeat reasoning above

What Can We Say for Arbitrary Priors?

- main result holds for a set of prior beliefs (characterized by the assumption)
- suppose DM has a prior outside of this set
- under optimal sampling, his posterior belief will eventually enter that set
- at that point the characterization again applies, so e.g.:

Corollary

Starting from any prior belief, the optimal information acquisition strategy is eventually a constant attention level proportional to the weight vector α.

Application 1:

Binary Choice

Binary Choice

- literature beginning with drift-diffusion model (Ratcliff, 1978)
  - two goods with unknown payoffs θ_1 and $-\theta_2$
  - agent can devote effort towards learning about these payoffs before making his decision
- DDM: agent's prior is supported on two values $\theta_L < \theta_H$, uncertainty is only over which good is better
- Fudenberg, Strack, and Strzalecki (2016): "uncertain-difference" DDM with $(\theta_1, -\theta_2) \sim \mathcal{N}(\mu, \Sigma)$
- result from FSS: assume $\Sigma = I$, then optimal attention choices constant at $(1/2, 1/2)$

Binary Choice

- this problem is nested in our setting as case of $\alpha_1 = \alpha_2 = 1$ (given which our characterization holds for all priors)

Binary Choice

- this problem is nested in our setting as case of $\alpha_1 = \alpha_2 = 1$ (given which our characterization holds for all priors)

Corollary

Starting from any prior with $\Sigma_{11} \geq \Sigma_{22}$, the DM first attends to attribute 1 exclusively, then switches to equal attention at time

$$t_1 = \frac{\Sigma_{11} - \Sigma_{22}}{\det(\Sigma)}.$$

Binary Choice

- this problem is nested in our setting as case of $\alpha_1 = \alpha_2 = 1$ (given which our characterization holds for all priors)

Corollary

Starting from any prior with $\Sigma_{11} \geq \Sigma_{22}$, the DM first attends to attribute 1 exclusively, then switches to equal attention at time

$$t_1 = \frac{\Sigma_{11} - \Sigma_{22}}{\det(\Sigma)}.$$

- generalizes the FSS result:
  - allows for correlation and asymmetry between unknown payoffs
  - applies "off-path" as well
- can use to derive comparative statics

Comparative Static in Initial Uncertainty

e.g. how does more initial uncertainty about an attribute affect the time path of attention?

Corollary

Suppose $\Sigma_{11} > \Sigma_{22}$ (more initial uncertainty about attribute 1).

1. *If $|\Sigma_{12}| < \Sigma_{22}$, increase in Σ_{11} leads to weakly higher attention to attribute 1 at every time.*
2. *Otherwise, increase in Σ_{11} leads to uniformly lower attention.*

Comparative Static in Initial Uncertainty

e.g. how does more initial uncertainty about an attribute affect the time path of attention?

Corollary

Suppose $\Sigma_{11} > \Sigma_{22}$ (more initial uncertainty about attribute 1).

1. *If $|\Sigma_{12}| < \Sigma_{22}$, increase in Σ_{11} leads to weakly higher attention to attribute 1 at every time.*
2. *Otherwise, increase in Σ_{11} leads to uniformly lower attention.*

- increasing initial uncertainty about attribute 1 changes the "switch point" between stages 1 and 2
- whether it moves it earlier or later depends on how correlated the attributes are

Intuition

suppose $|\Sigma_{12}|$ is small:

- then greater initial uncertainty about θ_1 increases the value to learning about it
- so increase in Σ_{11} results in more attention paid to attribute 1

Intuition

suppose $|\Sigma_{12}|$ is small:

- then greater initial uncertainty about θ_1 increases the value to learning about it
- so increase in Σ_{11} results in more attention paid to attribute 1

but large $|\Sigma_{12}|$ can reverse this:

- information about θ_1 also reveals about θ_2
- increasing Σ_{11} (for fixed Σ_{12}, Σ_{22}) *decreases correlation*, less externality
- faster for uncertainty about θ_1 to be reduced *relatively*
- this effect dominates when prior correlation is significant

Application 2:

Competing Information Providers

Competing Information Providers

- new sources have expertise on a topic (e.g. Mueller report), and provide information on this over time
- want to maximize time spent on their site
- choose the informativeness of news articles (i.e. reveal everything you know all at once vs. trickle it out slowly)
- in talk assume two sources, but see paper for extension to K sources

The Game

- $\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$
- payoff-relevant state $\theta_1 + \theta_2$

The Game

- $\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$

- payoff-relevant state $\theta_1 + \theta_2$

- each source $i = 1, 2$ freely chooses σ_i, providing

  $$\theta_i + \mathcal{N}(0, \sigma_i^2)$$

  per unit of time

- source i's payoff is the *discounted average attention*

  $$\int_0^\infty e^{-rt} \beta_i^t \, dt$$

The Game

- $\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$

- payoff-relevant state $\theta_1 + \theta_2$

- each source $i = 1, 2$ freely chooses σ_i, providing
  $$\theta_i + \mathcal{N}(0, \sigma_i^2)$$
  per unit of time

- source i's payoff is the *discounted average attention*
  $$\int_0^\infty e^{-rt} \beta_i^t \, dt$$

- note: not necessary to impose a cost to providing more precise information, equilibrium will have interior choices of σ_i

Equilibrium

Proposition

The unique equilibrium is a pure strategy equilibrium (σ^*, σ^*) *with*

$$\sigma^* = \sqrt{\frac{1-\rho}{2r}}$$

with ρ *being DM's prior correlation and* r *being the news sources' discount rate.*

- signals are more precise in equilibrium (lower σ^*) when news sources are less patient (larger r)

Role of Patience

$$\sigma^* = \sqrt{\frac{1-\rho}{2r}}.$$

increasing noise σ_i (i.e. provide lower-quality information) has two opposing effects on attention:

1. DM more likely to attend to other source initially
2. but in the long run, source i receives more attention: $\dfrac{\sigma_i}{\sigma_i + \sigma_j}$

Role of Patience

$$\sigma^* = \sqrt{\frac{1-\rho}{2r}}.$$

increasing noise σ_i (i.e. provide lower-quality information) has two opposing effects on attention:

1. DM more likely to attend to other source initially
2. but in the long run, source i receives more attention: $\dfrac{\sigma_i}{\sigma_i + \sigma_j}$

$\Longrightarrow$ if news sources are patient (small r), they provide noisy info

Role of Patience

$$\sigma^* = \sqrt{\frac{1-\rho}{2r}}.$$

increasing noise σ_i (i.e. provide lower-quality information) has two opposing effects on attention:

1. DM more likely to attend to other source initially
2. but in the long run, source i receives more attention: $\dfrac{\sigma_i}{\sigma_i + \sigma_j}$

$\Longrightarrow$ if news sources are patient (small r), they provide noisy info

$\Longrightarrow$ if news sources are impatient (large r), they compete to be chosen in stage 1

role of correlation

Conclusion

- we study the problem of dynamic allocation of attention across diverse information sources
- under condition on prior, solution is simple/tractable/robust
- useful towards various applications

Thank You!

Discrete-Time Analogue

Liang, Mu, and Syrgkanis (working paper):

- unknown attribute values $\theta_1, \ldots, \theta_K$ are jointly normal
- payoff-relevant state $\omega = \langle \alpha, \theta \rangle$ with a known and positive weight vector α
- at each discrete period t, agent chooses from among K information sources
- choice of source i produces observation of

$$Y_i = \theta_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}\left(0, \frac{1}{\Delta}\right)$$

Relationship Between Settings

- suppose in continuous-time model, DM's attention must be constant and degenerate over each of $[0, \Delta), [\Delta, 2\Delta)$, etc.
- the difference $X_i^{t+\Delta} - X_i^t$ is equivalent to the signal $\Delta \cdot Y_i$ in the discrete-time model
- taking $\Delta \to 0$ thus yields our main setting where attention choices can be changed continuously
- but in discrete-time, there is an "integer problem," since signals are non-divisible
- continuous-time formulation allows for a sharper characterization of the optimal info acquisition strategy, and conditions needed for this characterization to hold
- settings share an optimality of "myopic" acquisition

Counterexample

- unknown attributes
$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} 10 & -3 \\ -3 & 1 \end{pmatrix}\right),$$
want to learn $\theta_1 + 4\theta_2$

- at al times $t \leq 1/4$, t-optimal vector is $(t, 0)$

- for $t \in (1/4, 1]$, t-optimal vector is $\left(\frac{-t+1}{3}, \frac{4t-1}{3}\right)$

- thus as budget increases from $1/4$ to 1, optimal amount of attention devoted to θ_1 is *decreasing*

- so the t-optimal attention vectors are not monotone in t

Counterexample Intuition

- initially, marginal value of learning about θ_1 is strictly largest $\Rightarrow$ learn about θ_1
- at $t = 1/4$, marginal values have equalized
- turn from "first-order" comparison of marginal values to "second-order" comparison of mixtures between signals
- optimal mixture depends on whether the signals are substitutes or complements
- at $t = 1/4$, learning about θ_1 and θ_2 are *substitutes*
- information about attribute 1 has a large negative impact on the marginal value of information about attribute 2
- agent would optimally like to take away some attention from attribute 1 and re-distribute it to attribute 2

back

Transformation

Given σ_1, σ_2, we can normalize to unit signal precision:

- Define $\tilde{\theta}_i = \theta_i / \sigma_i$
- Then signal $\theta_i + \mathcal{N}(0, \sigma_i^2)$ is equivalent to $\tilde{\theta}_i + \mathcal{N}(0, 1)$, returns our model

Transformation

Given σ_1, σ_2, we can normalize to unit signal precision:

- Define $\tilde{\theta}_i = \theta_i / \sigma_i$
- Then signal $\theta_i + \mathcal{N}(0, \sigma_i^2)$ is equivalent to $\tilde{\theta}_i + \mathcal{N}(0,1)$, returns our model

- Payoff-relevant state rewritten as $\sigma_1 \tilde{\theta}_1 + \sigma_2 \tilde{\theta}_2$, so $\tilde{\alpha}_i = \sigma_i$
- Transformed prior covariance matrix of $\tilde{\theta}$ is

$$\tilde{\Sigma} = \begin{pmatrix} \frac{1}{\sigma_1^2} & \frac{\rho}{\sigma_1 \sigma_2} \\ \frac{\rho}{\sigma_1 \sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}$$

Condition on Prior Belief is Satistifed

Assumption satisfied since

$$\sigma_1 \left(\frac{1}{\sigma_1^2} + \frac{\rho}{\sigma_1 \sigma_2} \right) + \sigma_2 \left(\frac{\rho}{\sigma_1 \sigma_2} + \frac{1}{\sigma_2^2} \right) = (1 + \rho) \left(\frac{1}{\sigma_1} + \frac{1}{\sigma_2} \right) \geq 0.$$

Can thus use theorem to find attention levels given any σ_1, σ_2.

Role of Correlation

$$\sigma^* = \sqrt{\frac{1-\rho}{2r}}.$$

- if prior is negatively correlated (smaller ρ), signals are *complements*

  $\implies$ stage 1 is shorter

- thus more competition for the long run, and sources choose to provide noisier signals

back