Lecture 5 - Model Misspecification

Outline

- Topic: Model Misspecification
- Passive Learning (Berk 1966)
- Active Learning (Bohren & Hauser 2021 ECMA)
- Social Learning Frick, Iijima and Ishii (2020 ECMA)
- If time allows: my own work in measuring misspecification empirically

Framework

- Set of parameters/states/types: $\omega \in \Omega$
- Signals are iid: $X_1, ..., X_n \sim \ldots \overset{\text{iid}}{\sim} f(\cdot|\omega)$, f is DGP.
- Prior $p(\omega)$, full support

Framework

- Set of parameters/states/types: $\omega \in \Omega$
- Signals are iid: $X_1, ..., X_n \sim \ldots \overset{\text{iid}}{\sim} f(\cdot|\omega)$, f is DGP.
- Prior $p(\omega)$, full support
- Q: How does your belief converge after seeing large sequences?
  - It depends on your belief about $\hat{f}$ about f. If $\hat{f} = f$ then you converge to ω (assuming it is identifiable from f).

Framework

- Set of parameters/states/types: $\omega \in \Omega$
- Signals are iid: $X_1, ..., X_n \sim \ldots \overset{\text{iid}}{\sim} f(\cdot|\omega)$, f is DGP.
- Prior $p(\omega)$, full support
- Q: How does your belief converge after seeing large sequences?
  - It depends on your belief about $\hat{f}$ about f. If $\hat{f} = f$ then you converge to ω (assuming it is identifiable from f).
  - What if you do not? Berk (1966) shows you converge to the "best approximation" of ω given $\hat{f}$.

Convergence with correct specification

- If you had the correct belief then your posterior for ω after $\mathbf{X_k} = X_1, ..X_k$:

$$p(\omega|\mathbf{X_k}) = \frac{\prod_{i=1}^{k} f(X_i|\omega)p(\omega)}{\sum_{\omega}\prod_{i=1}^{k} f(X_i|\omega')p(\omega')}$$

Convergence with correct specification

- If you had the correct belief then your posterior for ω after $\mathbf{X_k} = X_1, ..X_k$:

$$p(\omega|\mathbf{X_k}) = \frac{\prod_{i=1}^{k} f(X_i|\omega)p(\omega)}{\sum_{\omega} \prod_{i=1}^{k} f(X_i|\omega')p(\omega')}$$

- Your posterior odds ratio between states ω and ω':

$$\frac{p(\omega|\mathbf{X_k})}{p(\omega'|\mathbf{X_k})} = \frac{\prod_{i=1}^{k} f(X_i|\omega)p(\omega)}{\prod_{i=1}^{k} f(X_i|\omega')p(\omega')}$$

Convergence with correct specification

- Q: As $k \to \infty$ if ω is the true state, what happens to:

$$\lim_{k\to\infty} \frac{p(\omega'|\mathbf{X_k})}{p(\omega|\mathbf{X_k})} = \frac{\prod_{i=1}^{k} f(X_i|\omega')p(\omega')}{\prod_{i=1}^{k} f(X_i|\omega)p(\omega)} = M_k$$

Convergence with correct specification

- Q: As $k \to \infty$ if ω is the true state, what happens to:

$$\lim_{k \to \infty} \frac{p(\omega' | \mathbf{X_k})}{p(\omega | \mathbf{X_k})} = \frac{\prod_{i=1}^{k} f(X_i | \omega') p(\omega')}{\prod_{i=1}^{k} f(X_i | \omega) p(\omega)} = M_k$$

- We will show the above converges to 0.

Convergence with correct specification

- As $n \to \infty$, the distribution of signals $(X_1, .., X_n)$ is equal to $f(X|\omega)$. Let signal types by $y_1, .., y_m$.

- Then by Glivenko-Cantelli (1933)

$$\lim_{k\to\infty} \frac{\prod_{i=1}^{k} f(X_i|\omega')p(\omega')}{\prod_{i=1}^{k} f(X_i|\omega)p(\omega)} = \lim_{N\to\infty} \frac{\prod_{i=1}^{m} f(y_i|\omega')^{Nf(y_i|\omega)}p(\omega')}{\prod_{i=1}^{m} f(y_i|\omega)^{Nf(y_i|\omega)}p(\omega)}$$

Convergence with correct specification

- As $n \to \infty$, the distribution of signals $(X_1, .., X_n)$ is equal to $f(X|\omega)$. Let signal types by $y_1, .., y_m$.

- Then by Glivenko-Cantelli (1933)

$$\lim_{k\to\infty} \frac{\prod_{i=1}^{k} f(X_i|\omega')p(\omega')}{\prod_{i=1}^{k} f(X_i|\omega)p(\omega)} = \lim_{N\to\infty} \frac{\prod_{i=1}^{m} f(y_i|\omega')^{Nf(y_i|\omega)}p(\omega')}{\prod_{i=1}^{m} f(y_i|\omega)^{Nf(y_i|\omega)}p(\omega)}$$

- Then take the log:

$$\lim_{N\to\infty} N\sum_{i=1}^{m} f(y_i|\omega)\log(f(y_i|\omega')) - N\sum_{i=1}^{m} f(y_i|\omega)\log(f(y_i|\omega)) + \log\frac{p(\omega')}{p(\omega)}$$

- Want to show the first term is smaller than the second, last term is finite.

Convergence with correct specification

- Suffice to see that for any $(y_1, .., y_n) \in [0,1]^n$ such that $\sum_i y_i = 1$:

$$(y_1, .., y_n) \in \arg\max_{x_i} \sum_i y_i \log(x_i)$$

such that $\sum_i x_i = 1$

Convergence with correct specification

- Suffice to see that for any $(y_1, .., y_n) \in [0,1]^n$ such that $\sum_i y_i = 1$:

$$(y_1, .., y_n) \in \arg\max_{x_i} \sum_i y_i \log(x_i)$$

  such that $\sum_i x_i = 1$

The Lagrangian is:

$$\mathcal{L}(x_1, .., x_n, \lambda) = \sum_i y_i \log(x_i) + (1 - \sum_i x_i)$$

Partial gives: $x_j = \frac{y_j}{\lambda}$
We have then $\sum_j \frac{y_j}{\lambda} = 1$ so $\lambda = 1$.

Proof (via KL divergence)

1. Rewrite the objective:

$$\sum_{i=1}^{n} y_i \log(x_i) \;=\; -\sum_{i=1}^{n} y_i \log\!\left(\frac{y_i}{x_i}\right) \;+\; \sum_{i=1}^{n} y_i \log(y_i) \;=\; -\,D_{\mathrm{KL}}(y \,\|\, x) \;+\; \text{constant in } x.$$

2. Minimizing $D_{\mathrm{KL}}(y \,\|\, x) = \sum_{i=1}^{n} y_i \log\big(y_i/x_i\big)$ over x on the simplex uniquely forces $x = y$ (KL divergence is nonnegative and is zero *if and only if* $x = y$).
3. Hence $x = y$ is exactly the maximizer of $\sum_i y_i \log(x_i)$.

Convergence with misspecification

- Take any two states ω and ω', can rewrite, and now it's just $\hat{f}$:

$$\lim_{k\to\infty} \frac{p(\omega'|\mathbf{X_k})}{p(\omega|\mathbf{X_k})} = \frac{\prod_{i=1}^{k} \hat{f}(X_i|\omega')p(\omega')}{\prod_{i=1}^{k} \hat{f}(X_i|\omega)p(\omega)}$$

Convergence with misspecification

- Take any two states ω and ω', can rewrite, and now it's just $\hat{f}$:

$$\lim_{k\to\infty} \frac{p(\omega'|\mathbf{X_k})}{p(\omega|\mathbf{X_k})} = \frac{\prod_{i=1}^{k} \hat{f}(X_i|\omega')p(\omega')}{\prod_{i=1}^{k} \hat{f}(X_i|\omega)p(\omega)}$$

- Now suppose the true state is ω^*, recall y_i are signal realization types:

$$\lim_{k\to\infty} \frac{\prod_{i=1}^{k} \hat{f}(X_i|\omega')p(\omega')}{\prod_{i=1}^{k} \hat{f}(X_i|\omega)p(\omega)} = \lim_{N\to\infty} \frac{\prod_{i=1}^{m} \hat{f}(y_i|\omega')^{Nf(y_i|\omega^*)}p(\omega')}{\prod_{i=1}^{m} \hat{f}(y_i|\omega)^{Nf(y_i|\omega^*)}p(\omega)}$$

Convergence with misspecification

- Take any two states ω and ω', can rewrite, and now it's just $\hat{f}$:

$$\lim_{k\to\infty} \frac{p(\omega'|\mathbf{X_k})}{p(\omega|\mathbf{X_k})} = \frac{\prod_{i=1}^{k} \hat{f}(X_i|\omega')p(\omega')}{\prod_{i=1}^{k} \hat{f}(X_i|\omega)p(\omega)}$$

- Now suppose the true state is ω^*, recall y_i are signal realization types:

$$\lim_{k\to\infty} \frac{\prod_{i=1}^{k} \hat{f}(X_i|\omega')p(\omega')}{\prod_{i=1}^{k} \hat{f}(X_i|\omega)p(\omega)} = \lim_{N\to\infty} \frac{\prod_{i=1}^{m} \hat{f}(y_i|\omega')^{Nf(y_i|\omega^*)}p(\omega')}{\prod_{i=1}^{m} \hat{f}(y_i|\omega)^{Nf(y_i|\omega^*)}p(\omega)}$$

- Then the ω which maximizes $\hat{f}(y_i|\omega)^{Nf(y_i|\omega^*)}$ will be the convergent one (if there is more than 1, then you converge to prior ratio).

Convergence with misspecification

- Note that maximizing $\hat{f}(y_i|\omega)^{N f(y_i|\omega^*)}$ is the same as minimizing $D_{\mathrm{KL}}(f(\cdot|\omega^*) \,\|\, \hat{f}(\cdot|\omega))$

$$\sum_{i=1}^{n} f(y_i|\omega^*)\, \log(\hat{f}(y_i|\omega)) \;=\; -\sum_{i=1}^{n} f(y_i|\omega^*)\, \log\Big(\frac{f(y_i|\omega^*)}{\hat{f}(y_i|\omega)}\Big) + \sum_{i=1}^{n} f(y_i|\omega^*)\, \log(f(y_i|\omega^*))$$

$$= \; - D_{\mathrm{KL}}(f(y_i|\omega^*) \,\|\, \hat{f}(y_i|\omega)) \; + \; \text{constant in } x.$$

Berk 1966

Berk 1966 (imprecise form)

Given $f, \hat{f}$ and true state ω, the belief converges to ω^* such that

$$\omega^* = \arg\min_{\omega} D_{\mathrm{KL}}(f(\cdot|\omega^*) \,\|\, \hat{f}(\cdot|\omega))$$

Berk 1966

Berk 1966 (imprecise form)

Given $f, \hat{f}$ and true state ω, the belief converges to ω^* such that

$$\omega^* = \arg\min_{\omega} D_{\mathrm{KL}}(f(\cdot|\omega^*) \parallel \hat{f}(\cdot|\omega))$$

Your belief converges to the state that best fits the data (given your misspecified model).

Example

- Unknown state $\omega \in \{L, R\}$.
- Each period, choose one of two platform $a_t \in \{L, R\}$ and receive a signal.
- $u(a_t, \omega) = 1$ if and only if $a_t = \omega$.
- For both actions: $f(r|\omega) = \exp(-\beta_\omega)$ with $0 < \beta_R < \beta_L$ and $f(l|\omega) = 1 - \exp(-\beta_\omega)$
  Agent misestimates the probability of news r by a factor k:
  - $\hat{f}^\omega(r) = \exp(-\beta_\omega - k))$ with $k \in (-\infty, \beta_R)$

- Let $\gamma(\omega)$ be difference between the KLs (of true and misspecified) of the two states given true state ω.

$$\begin{aligned}\gamma(\omega) &= D_{\mathrm{KL}}(f^{\omega} \,\|\, \hat{f}^{L}) - D_{\mathrm{KL}}(f^{\omega} \,\|\, \hat{f}^{R}) \\ &= \exp(-\beta_\omega) \log \frac{\exp(-\beta_R - k)}{\exp(-\beta_L - k)} + (1 - \exp(-\beta_\omega)) \log \frac{1 - \exp(-(\beta_R - k))}{1 - \exp(-(\beta_L - k))}\end{aligned}$$

- Let $\gamma(\omega)$ be difference between the KLs (of true and misspecified) of the two states given true state ω.

$$\begin{aligned}\gamma(\omega) &= D_{\mathrm{KL}}(f^\omega \parallel \hat{f}^L) - D_{\mathrm{KL}}(f^\omega \parallel \hat{f}^R) \\ &= \exp(-\beta_\omega)\log\frac{\exp(-\beta_R - k)}{\exp(-\beta_L - k)} + (1-\exp(-\beta_\omega))\log\frac{1-\exp(-(\beta_R - k))}{1-\exp(-(\beta_L - k))}\end{aligned}$$

- Then if $\gamma(\omega) > 0$ converges to L otherwise L.

- Let $\gamma(\omega)$ be difference between the KLs (of true and misspecified) of the two states given true state ω.

$$\begin{aligned}
\gamma(\omega) &= D_{\mathrm{KL}}(f^{\omega} \parallel \hat{f}^{L}) - D_{\mathrm{KL}}(f^{\omega} \parallel \hat{f}^{R}) \\
&= \exp(-\beta_{\omega}) \log \frac{\exp(-\beta_R - k)}{\exp(-\beta_L - k)} + (1 - \exp(-\beta_{\omega})) \log \frac{1 - \exp(-(\beta_R - k))}{1 - \exp(-(\beta_L - k))}
\end{aligned}$$

- Then if $\gamma(\omega) > 0$ converges to L otherwise L.
- let $(\beta_L, \beta_L) = (10, 2)$ then:
  - Correct in L and incorrect in R for $K < -0.55$
  - Correct in both states when $k \in (-0.55, 0.41)$
  - Incorrect in L and correct in R for $k > 0.41$

Active Learning

- So far, only considered passive learning - information not dependent on action chosen.
- Many economic environments feature instead active learning.

Active Learning

- So far, only considered passive learning - information not dependent on action chosen.
- Many economic environments feature instead active learning.
- Ex-ante, seems tricky, because different beliefs lead to different actions.
    - Asymptotic properties may depend on how signal distribution varies with beliefs (since action chosen depends on beliefs)

Active Learning

- So far, only considered passive learning - information not dependent on action chosen.
- Many economic environments feature instead active learning.
- Ex-ante, seems tricky, because different beliefs lead to different actions.
  - Asymptotic properties may depend on how signal distribution varies with beliefs (since action chosen depends on beliefs)

Bohren and Hauser (2021 ECMA) show that under fairly mild conditions:

- Only need to consider model fit at optimal actions for degenerate beliefs.
- We will briefly go over their result in the binary case.

Definition

A belief $p \in \{0,1\}$ is locally stable if there exists an $\epsilon > 0$ such that if $||p_1 - p|| < \epsilon$ then $\Pr(p_t \to p) > 0$ and is globally stable if for any interior $p_1 \in (0,1)$, $\Pr(p_t \to p) > 0$.

- For another belief (close or not close), the optimal actions leads to a non-zero probability of converging to $p \in \{0,1\}$.
- Note stability is NOT about correct convergence.
- When is a belief locally or globally stable?

- Suppose a belief is locally stable. It is also globally stable is the local set is reached with positive probability.

- Now for global stability, we define:

$$\omega^*(\omega, a) = \arg \min_{\omega' \in \{L,R\}} D_{\mathrm{KL}}(f^{\omega}(.;a) \,\|\, \hat{f}^{\omega'}(.;a))$$

- Define also $a^*(p)$ the optimal action at p.

- Observation (under mild identification assumption): Given the true state ω, $p = 0$ is globally stable if and only if $\omega^*(\omega, a^*(0)) = \{L\}$ and $p = 0$ is globally stable if and only if $\omega^*(\omega, a^*(q)) = \{R\}$. No belief $p \in (0, 1)$ is locally or globally stable.

- For intuition: Think of the two signal case.

Theorem

Consider a myopic agent. WHen the the true state is L then :

1. *Learning is almost surely correct iff* $\omega^*(\omega, a(p)) = \{L\}$ *for* $p \in \{0, 1\}$.
2. *Learning is almost surely incorrect iff* $\omega^*(\omega, a(p)) = \{R\}$ *for* $p \in \{0, 1\}$.
3. *Learning is path dependent with positive probability of correct and incorrect iff* $\omega^*(\omega, a(0)) = \{L\}$ *and* $\omega^*(\omega, a(1)) = \{R\}$.
4. *LLearning is almost surely cyclical if* $\omega^*(\omega, a(0)) = \{R\}$ *and* $\omega^*(\omega, a(1)) = \{L\}$.

Example

- Consider our news example with now active learning:
  - $f^\omega(r, R) = \exp(-\beta_\omega + k)$ on platform R.
  - $f^\omega(r, L) = \exp(-\beta_\omega - k)$ on platform L.
  - Agent fails to understand this: $\hat{f}^\omega(r, L) = \hat{f}^\omega(r, R) = \exp(-\beta_\omega + \hat{k})$.

Example

- As before, we compute $\gamma(\omega, a^*(0))$, at $p = 0$, the choose platform L, so $a^*(0) = L$.

$$\gamma(\omega; a^(0)) = \exp(-\beta_L - k)\,\log\left(\frac{\exp(-\beta_L - k)}{\exp(-\beta_L) - k}\right) + \big(1 - \exp(-\beta_L - k)\big)\,\log\left(\frac{1 - \exp(-\beta_L - k)}{1 - \exp(-\beta_L) - k}\right).$$

Example

- As before, we compute $\gamma(\omega, a^*(0))$, at $p = 0$, the choose platform L, so $a^*(0) = L$.

$$\gamma(\omega; a^(0)) = \exp(-\beta_L - k)\ \log\left(\frac{\exp(-\beta_L - k)}{\exp(-\beta_L) - k}\right) + \left(1 - \exp(-\beta_L - k)\right)\ \log\left(\frac{1 - \exp(-\beta_L - k)}{1 - \exp(-\beta_L) - k}\right).$$

- Similarly for $p = 1$ we have can compute $\gamma(\omega, a^*(1))$.

- Can check that $\gamma(\omega; a^(0))$ is increasing in $\hat{k}$ while $\gamma(\omega, a^*(1))$ is decreasing.

Example

- As before, we compute $\gamma(\omega, a^*(0))$, at $p = 0$, the choose platform L, so $a^*(0) = L$.

$$\gamma(\omega; a^(0)) = \exp(-\beta_L - k) \log\left(\frac{\exp(-\beta_L - k)}{\exp(-\beta_L) - k}\right) + (1 - \exp(-\beta_L - k)) \log\left(\frac{1 - \exp(-\beta_L - k)}{1 - \exp(-\beta_L) - k}\right).$$

- Similarly for $p = 1$ we have can compute $\gamma(\omega, a^*(1))$.
- Can check that $\gamma(\omega; a^(0))$ is increasing in $\hat{k}$ while $\gamma(\omega, a^*(1))$ is decreasing.
- Then if $\hat{k}$ is far from k, $\gamma(\omega, a^*(0)) < 0$ and $\gamma(\omega, a^*(0)) > 0$ (what's the interpretation?).

Robustness

Robustness

Observation 4 (Robustness of Learning). Let $m_1 := \big(f_1^L(\cdot;\omega),\, f_1^R(\cdot;\omega)\big)_{\omega\in A}$ be an identified subjective model such that

$$\gamma(\omega;a) \neq 0 \quad \text{for} \quad a \in \{a^*(0),\, a^*(1)\} \quad \text{and} \quad \omega \in \{L,\, R\}.$$

Then there exists $\alpha > 0$ such that any identified subjective model

$$m_2 := \big(f_2^L(\cdot;\omega),\, f_2^R(\cdot;\omega)\big)_{\omega\in A}$$

that is sufficiently close to m_1 in terms of KL divergence,

$$D_{\mathrm{KL}}\big(f_2^i(\cdot;\omega)\,\big\|\,f_1^i(\cdot;\omega)\big) < \alpha \quad \text{for} \quad \omega \in \{a^*(0),\, a^*(1)\} \quad \text{and} \quad \omega \in \{L,\, R\},$$

has the same learning outcomes.

Frick, Iijima and Ishii (2020 ECMA)

- Setting: We know information aggregation works in general when agents are Bayesian, well-specified, and observe sufficient information.

- Q: What happens when agents are misspecified about the information structure?

Frick, Iijima and Ishii (2020 ECMA)

- Setting: We know information aggregation works in general when agents are Bayesian, well-specified, and observe sufficient information.
- Q: What happens when agents are misspecified about the information structure?
- One guess: It depends on the amount of misspecification... if little maybe aggregate is only a little off.

Frick, Iijima and Ishii (2020 ECMA)

- Setting: We know information aggregation works in general when agents are Bayesian, well-specified, and observe sufficient information.
- Q: What happens when agents are misspecified about the information structure?
- One guess: It depends on the amount of misspecification... if little maybe aggregate is only a little off.
- They show: arbitrarily small misspecification can lead to full failure, misspecification is exacerbated by social learning.

Environment

- Time period $t = 0, 1, 2, 3...$
- Continuum of agents with type $\theta \in \mathbb{R}$ - private information.
- Type $\theta \sim F$.
- State of the world is $\omega \sim \Psi$, $\omega \in [\underline{\omega}, \overline{\omega}]$, draw at $t = 0$
- Agent i picks $a_{it} \in \{0, 1\}$ each period to myopically maximize his EU given belief about ω
- Utility is $u(a, \theta, \omega)$

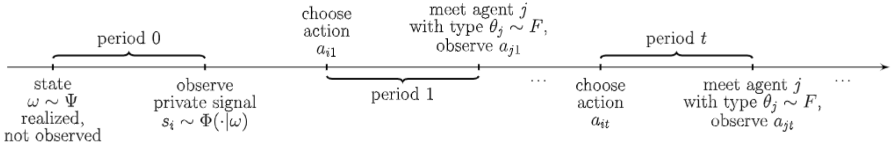
Information

- At $t = 0$, i observes $s_i \in \mathbb{R}$, iid. Assume monotone likelihood ratio property, higher states more likely to draw higher signals
- At each period, each agent i random meets another agent j and observes a_{jt}
- Then at each period, agents have s_i plus all observed actions as information

Information

- At $t = 0$, i observes $s_i \in \mathbb{R}$, iid. Assume monotone likelihood ratio property, higher states more likely to draw higher signals
- At each period, each agent i random meets another agent j and observes a_{jt}
- Then at each period, agents have s_i plus all observed actions as information
- Crucially: agents do not observe their utility - otherwise immediately learn the state
  - Interpretation: payoff only realize in the long run, or these are successive shortlived agents

Timeline


period 0
choose
action
a_{i1}
meet agent j
with type $\theta_j \sim F$,
observe a_{j1}
period t
state
$\omega \sim \Psi$
realized,
not observed
observe
private signal
$s_i \sim \Phi(\cdot|\omega)$
period 1
...
choose
action
a_{it}
meet agent j
with type $\theta_j \sim F$,
observe a_{jt}
...

Threshold Action

- Important feature: given any ω and F, at time period 1, there are exactly $q(\omega, F)$ many agents who choose action 1, this is strictly increasing in ω.
- Denote by $\theta^*(\omega)$ the threshold type for whom they are indifference between $a = 1$ and $a = 0$ given state ω.

Some Benchmarks

- Let $\hat{F}$ be the perceived distribution of types.

Some Benchmarks

- Let $\hat{F}$ be the perceived distribution of types.

Lemma

Suppose $\hat{F} = F$ then in any state ω, almost all agents' beliefs converge to a point mass on ω.

Some Benchmarks

- Let $\hat{F}$ be the perceived distribution of types.

Lemma

Suppose $\hat{F} = F$ then in any state ω, almost all agents' beliefs converge to a point mass on ω.

Actually not trivial due to the dynamics of social learning.

- If agents only pick their action once - then you converge as you see the distribution of actions which allows you to infer the distribution of signals - given you know the distribution of types.
- Social learning makes this argument harder but then main result goes through.

Misperception

- Let $\hat{F}$ be the perceived distribution of types and now let's consider actions are fixed from the start. So here only misspecification and no social learning.

Lemma

Given a true state ω^*, agents' long run belief assigns probability 1 to the state $\hat{\omega} = \arg\min_\omega D_{\mathrm{KL}}(q(\omega^*, F) \| \, q(\omega, \hat{F}))$. Where $q(\omega, F)$ is the distribution of actions given the signal structure, ω and F.

Misperception

- Let $\hat{F}$ be the perceived distribution of types and now let's consider actions are fixed from the start. So here only misspecification and no social learning.

Lemma

Given a true state ω^*, agents' long run belief assigns probability 1 to the state $\hat{\omega} = \arg\min_\omega D_{\mathrm{KL}}(q(\omega^*, F) \| \ q(\omega, \hat{F}))$. Where $q(\omega, F)$ is the distribution of actions given the signal structure, ω and F.

This is essentially straightforward from Berk.

Misperception

- Let $\hat{F}$ be the perceived distribution of types and now let's consider actions are fixed from the start. So here only misspecification and no social learning.

Lemma

Given a true state ω^*, agents' long run belief assigns probability 1 to the state $\hat{\omega} = \arg\min_{\omega} D_{\mathrm{KL}}(q(\omega^*, F)\| \, q(\omega, \hat{F}))$. Where $q(\omega, F)$ is the distribution of actions given the signal structure, ω and F.

This is essentially straightforward from Berk. So if actions are fixed at the start, and there is no dynamic social learning, the effect of misspecification is small whenever misspecification is small.

Breakdown of Learning

Theorem

Fix any F and $\hat{\omega}$. For any $\epsilon > 0$, there exists $\hat{F}$ such that $\| F - \hat{F}\| < \epsilon$ under which, in any state ω, almost all agents' beliefs converge to a point mass on $\hat{\omega}$.

Breakdown of Learning

Theorem

Fix any F and $\hat{\omega}$. For any $\epsilon > 0$, there exists $\hat{F}$ such that $\| F - \hat{F} \| < \epsilon$ under which, in any state ω, almost all agents' beliefs converge to a point mass on $\hat{\omega}$.

For any prior F and any state $\hat{\omega}$, there is an ϵ close misperception such that no matter what the true state is, people converge to it.

Proof: Step 1

- Consider a model where agents observe ALL other agents' actions each period and updating is slightly different.
- Then given some $\hat{F}$, what happens at period $t = 1$?

Proof: Step 1

- Consider a model where agents observe ALL other agents' actions each period and updating is slightly different.

- Then given some $\hat{F}$, what happens at period $t = 1$?
  - Belief converges to $\hat{\omega}_1 = \arg\min_\omega D_{\mathrm{KL}}(q(\omega^*, F) \| \, q(\omega, \hat{F}))$.

Proof: Step 1

- Consider a model where agents observe ALL other agents' actions each period and updating is slightly different.
- Then given some $\hat{F}$, what happens at period $t = 1$?
  - Belief converges to $\hat{\omega}_1 = \arg\min_\omega D_{\text{KL}}(q(\omega^*, F) \| \, q(\omega, \hat{F}))$.
- Then at time 2, all agents pick a as if $\hat{\omega}$ was the true state. This generates a distribution of action $q(\hat{\omega}_1, \hat{F})$ which is different from $q(\hat{\omega}_1, F)$. So at time 2, what's the new belief?

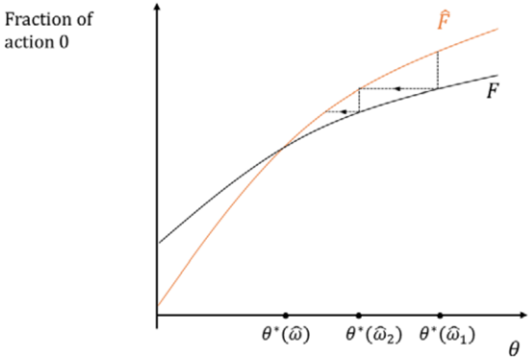
Proof: Step 1

- Consider a model where agents observe ALL other agents' actions each period and updating is slightly different.

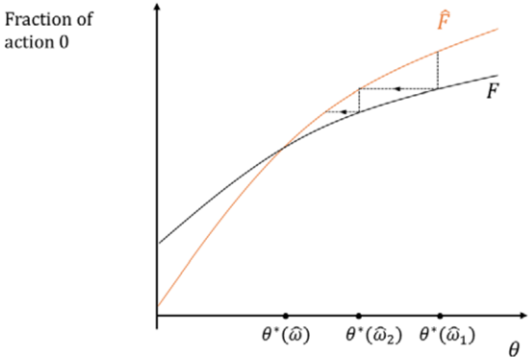
- Then given some $\hat{F}$, what happens at period $t = 1$?
  - Belief converges to $\hat{\omega}_1 = \arg\min_\omega D_{\mathrm{KL}}(q(\omega^*, F) \| \, q(\omega, \hat{F}))$.

- Then at time 2, all agents pick a as if $\hat{\omega}$ was the true state. This generates a distribution of action $q(\hat{\omega}_1, \hat{F})$ which is different from $q(\hat{\omega}_1, F)$. So at time 2, what's the new belief?
  - Belief converges to $\hat{\omega}_2 = \arg\min_\omega D_{\mathrm{KL}}(q(\hat{\omega}_1, F) \| \, q(\omega, \hat{F}))$.

Graphical Illustration



Graphical Illustration



No matter that the true F, ω are, can find some $\hat{F}$ that only intersects once with F and the beliefs must converge to the point of intersection.