Learning from cycles

Mehmet Dalkır

April 30, 2007

preliminary copy

Abstract

In this paper, Bayesian learning rules with distributions on non-Euclidean spaces are discussed. The findings of this study is crucial for the cases when individuals are forecasting a cyclic variable and learning from a noisy cyclic signal, like the index traders in the stock market. Our findings show that for a signal with a fixed precision for sale, an information provider has incentives from the public expectation that their signal value will be close to the mean of the public belief, in which case also from the prior public belief to have high precision. We make comments about the incentive of such providers to send spam e-mail messages, and also about the incentive to fabricate proofs supporting the common belief.

In a dynamic framework where agents learn from a signal and from a stochastic model, the linear and circular cases are found to be fundamentally similar.

1. Introduction

Coverage of nonlinear functional forms in economic and econometric theory is extensive.[1] The complexity of economic models may not be necessarily caused by the functional form, but by the spaces that the economic variables belong to. One aim of this paper is to explore the effect of migrating economic variables from the extensively used Euclidean spaces (namely, $\mathbb{R}^k$) to non-Euclidean spaces like a circle.

Bayesian social belief systems are usually based on distributions with supports in some Euclidean space, mainly the line. This study concentrates on Bayesian learning when the distributions have non-Euclidean support, like a circle or a torus.

The circular approach to Bayesian learning becomes crucial when we are dealing with cyclic variables rather than linear. Some examples are the macroeconomic aggregates (interest rates, exchange rates, price level, industrial production) and the stock market indices. As opposed to the linear case, the variable under consideration in the circular case is the phase, or stage of the variable rather than its level. For a procedure to identify the phase of the business cycle, see Harding and Pagan (2006). There are certain gains in terms of computation and identification with the phase approach.

One obvious gain is the possibility to define information which is "assertive," or "orthogonal," or "negative" relative to the public belief. When working with distributions on the line with infinite support, identifying these relations between signals and beliefs may not be possible. A good example is the widely used Gaussian learning rule, where the prior and posterior beliefs, as well as the information bearing signal are all random variables distributed with the Gaussian density. In Gaussian learning rule, *any* information with non-zero precision contributes to the precision of the prior belief. With our alternative setting, a "negative" signal may well decrease the precision of the public belief.

2. The linear case

This discussion closely follows the extensive review by Chamley (2004). The information

[1] For a review, see Barnett et al. 1989.

structure of the model is as follows:

1. The value of the nature's parameter θ is chosen randomly before the first period according to a normal distribution $N(\bar{\theta}, \sigma_\theta^2)$.

2. There is a countable number n of individuals that receive a private signal s_i, $i = 1, \ldots, n$: $s_i = \theta + \epsilon_i$. $Corr(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$, and $\epsilon \sim N(0, \sigma_\epsilon^2)$. All individuals have the same payoff function from their actions a: $U(a) = -E[(a - \theta)^2]$. Individual t chooses her action $a_t \in \mathbb{R}$ once and for all in period t. Each individual chooses an action only once, and at each time period, there is only one individual acting. The individual i is assigned to time t exogenously.

3. The public information at the beginning of period t is made of the prior distribution $N(\bar{\theta}, \sigma_\theta^2)$ and the set of previous actions $H_t = \{a_{t-1}, \ldots, a_1\}$.

2.1 Perfect observation of actions

The common quadratic payoff function induces the individuals to choose their beliefs $\tilde{\mu}_t \triangleq E_t[\theta]$ as their action: $a_t = \tilde{\mu}_t$ for all t. Since all individuals know the common payoff function, and aware that it is at the same time common information, they are aware that the actions reflect beliefs perfectly. So the public belief one period later, μ_{t+1} is equal to the individual t's personal belief: $\mu_{t+1} = \tilde{\mu}_t$. The public belief $\beta_t \sim N(\mu_t, \sigma_t^2)$ is updated according to the Bayesian rule:

$$\mu_{t+1} = (1 - \alpha_t)\mu_t + \alpha_t s_t, \tag{1}$$

$$\rho_{t+1} = \rho_t + \rho_\epsilon. \tag{2}$$

where $\alpha_t = \frac{\rho_\epsilon}{\rho_\epsilon + \rho_t}$, and $\rho_z \triangleq \sigma_z^{-2}$ is the precision of a random variable Z in general. As shorthand, $\mu_t \triangleq \mu_{\beta_t}$ and $\rho_t \triangleq \rho_{\beta_t}$ are used.

The updating of mean and variance in (1) and in (2) of the public belief is referred to as "Gaussian learning rule" and is a consequence of the normal distribution being the *conjugate* *family* of itself. [2] The Gaussian updating rule is particularly interesting since the updating of the precision ρ is independent of the mean μ.

In the linear case, the recursive equation (2) could be identified as a function of time:

$$\rho_{t+1} = \rho_\theta + t\rho_\epsilon. \tag{3}$$

So precision grows indefinitely in time, and public belief becomes "almost certain" as $t \to \infty$. The variance $\sigma_t^2 = 1/\rho_t$ converges to zero like $1/t$.

3. The circular case

The counterpart of the normal distribution on the circle is the von Mises distribution which has some desirable properties similar to the normal distribution. These properties are given by Mardia and Jupp (2000, pp.41-43). In this study, one additional property of the von Mises distribution is crucial: like the normal distribution, it is the conjugate family of itself. That is, if the prior (distribution of β_t) and the signal x_t are von Mises, then the posterior (distribution of β_{t+1}) is von Mises as well.

The setup of the model is similar to the linear case:

1. The value of the nature's parameter μ is chosen randomly before the first period according to a von Mises distribution $M(\delta, \kappa_\mu)$, where δ is the *mean direction* and κ_μ is the precision.

2. There is a countable number n of individuals that receive a private signal x_i, where the stochastic process $\{x_i\}_{i=1,\ldots,n}$ is defined on a probability space $(\Omega, \mathcal{F}, P)$ where $\mathcal{F}$ is a σ-algebra on the set Ω, and P is a probability measure on the measurable space $(\Omega, \mathcal{F})$. Specifically, $x_i = \mu + \epsilon_i$. $Corr(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$, and $\epsilon \sim M(0, \kappa)$. All individuals have the same payoff function from their actions a: $U(a) = -E[(a-\mu)^2]$. Individual t chooses his action $a_t \in \mathbb{R}$ once and for all in period t. Each individual chooses an

[2] That is, if the signal s_t is normal, then the set of possible common distributions for the prior β_t and the posterior β_{t+1} form the conjugate family of the normal distribution.

action only once, and at each time period, there is only one individual acting. The individual i is assigned to time t exogenously.

3. The public information at the beginning of period t is made of the prior distribution $M(\delta, \kappa_\mu)$ and the set of previous actions $H_t = \{a_{t-1}, \ldots, a_1\}$.

3.1 Perfect observation of actions

The common payoff function induces the individuals to choose their beliefs $\tilde{\mu}_t \triangleq E_t[\mu]$ as their action: $a_t = \tilde{\mu}_t$ for all t. Since all individuals know the payoff function is common, they are aware that the actions reflect beliefs perfectly. So the public belief one period later, μ_{t+1} is equal to the individual t's personal belief: $\mu_{t+1} = \tilde{\mu}_t$. The public belief $\beta_t \sim M(\mu_t, \kappa_t)$ is updated according to the Bayesian rule below (Bagchi, 1994) to become the public belief for the next period: $\beta_{t+1} \sim M(\mu_{t+1}, \kappa_{t+1})$.

$$
\begin{aligned}
\mu_{t+1} &= \tan^{-1} \frac{\kappa \sin x_t + \kappa_t \sin \mu_t}{\kappa \cos x_t + \kappa_t \cos \mu_t}, & (4) \\
\kappa_{t+1}^2 &= \kappa_t^2 + \kappa^2 + 2\kappa\kappa_t \cos(\mu_t - x_t). & (5)
\end{aligned}
$$

The updating rule in the von Mises case is particularly interesting: unlike the normal case, the updating rule for precision depends on the signal x_t and the prior mean μ_t. Three extreme cases will be considered:

1. *Assertive signal*

   When $x_t = \mu_t$ for a certain time interval t, then κ_t grows by the precision of the signal: $\kappa_{t+1} = \kappa_t + \kappa$. This case is identical with the Gaussian case.

2. *Orthogonal signal*

   When $|\mu_t - x_t| = \pi/2$ at a certain time interval t, then κ_t^2 grows by κ^2: $\kappa_{t+1}^2 = \kappa_t^2 + \kappa^2 \Rightarrow \kappa_{t+1} < \kappa_t + \kappa$.

3. *Negative signal*

When $|\mu_t - x_t| = \pi$ at a certain time interval t, then κ_t may increase or decrease: $\kappa_{t+1} = |\kappa_t - \kappa|$.

In general, $\lim_{t\to\infty} P(A(\kappa)\kappa t < \kappa_{t+1} < \kappa t) = 1$.[3] If the incoming signal has any vector component in the direction of the common belief, definitely there will be gains in terms of precision. Specifically, the following proposition applies:

Proposition 1 *$\kappa_{t+1} > \kappa_t$ when $\cos(\mu_t - x_t) > -\frac{\kappa}{2\kappa_t}$. If κ_t is growing, history is gaining more importance over the private signal through formula (4).*

Next, we will propose and prove a lemma showing that κ_{t+1}/κ converges to $tA(\kappa)$ in distribution.

Lemma 1 *$\lim_{t\to\infty} P(\kappa_{t+1}/\kappa > tA(\kappa)) = 1$.*

Proof.

From (4) and (5) we have the recursive relation

$$z_{t+1} = z_t + u_t \Rightarrow z_{t+1} = z_o + \sum_{j=1}^{t} u_j. \tag{6}$$

for z_t and u_t complex where $z_t \equiv \kappa_t \cos\mu_t + i\kappa_t \sin\mu_t$, $u_t \equiv \kappa\cos x_t + i\kappa \sin x_t$, and $z_o \equiv \kappa_\mu \cos\delta + i\kappa_\mu \sin\delta$.

Then we obtain the real recursive equations

$$\kappa_{t+1}\cos\mu_{t+1} = \kappa_\mu \cos\delta + \kappa\sum_{j=1}^{t}\cos x_j \quad\Rightarrow\quad C \triangleq \sum_{j=1}^{t}\cos x_j = \frac{\kappa_{t+1}}{\kappa}\cos\mu_{t+1} - \frac{\kappa_\mu}{\kappa}\cos\delta, \tag{7}$$

$$\kappa_{t+1}\sin\mu_{t+1} = \kappa_\mu \sin\delta + \kappa\sum_{j=1}^{t}\sin x_j \quad\Rightarrow\quad S \triangleq \sum_{j=1}^{t}\sin x_j = \frac{\kappa_{t+1}}{\kappa}\sin\mu_{t+1} - \frac{\kappa_\mu}{\kappa}\sin\delta. \tag{8}$$

[3]see proof of Lemma 1. $0 < A(\kappa) \triangleq I_1(\kappa)/I_o(\kappa) < 1$, for $0 < \kappa < \infty$ where $I_q(\cdot)$ is the modified Bessel function of the first kind of order q.

C and S are the sufficient statistics that represent the history H_t similar to the sufficient statistic $\sum_{j=1}^{t} s_j$ in the linear case. Mardia and Jupp (2000) give the distributions for $R \equiv \sqrt{C^2 + S^2}$, and $\bar{\theta} \equiv \tan^{-1}(S/C)$. We will establish divergence of κ_{t+1}/κ without using the distribution function explicitly. We use (7) and (8) and $R^2 = C^2 + S^2$ to derive an expression for $\kappa^2 R^2$:

$$\kappa^2 R^2 = \kappa_{t+1}^2 + \kappa_\mu^2 - 2\kappa_{t+1}\kappa_\mu \cos(\mu_{t+1} - \delta). \tag{9}$$

From Mardia and Jupp (2000, p.75)

$$E[\kappa^2 R^2] = \kappa^2 t^2 A^2(\kappa) + \kappa^2 t[1 - A^2(\kappa)] = \kappa_\mu^2 + E[\kappa_{t+1}] - 2\kappa_\mu E[\kappa_{t+1}\cos(\mu_{t+1} - \delta)]. \tag{10}$$

Collecting terms and dividing by t^2, we get

$$E[\kappa_{t+1}/t^2] = \kappa^2 A^2(\kappa) + \frac{\kappa^2}{t}[1 - A^2(\kappa)] - \frac{\kappa_\mu^2}{t^2} + 2\kappa_\mu E[\frac{1}{t^2}\kappa_{t+1}\cos(\mu_{t+1} - \delta)]. \tag{11}$$

All the terms except the first vanishes as $t \to \infty$. The last term vanishes since from the case of an assertive signal we know $\sup_\omega \kappa_{t+1}(\omega) = \kappa_\mu + \kappa t$, $\forall \omega \in \Omega$ and thus $\lim_{t\to\infty} \kappa_{t+1}/t^2 = 0$ with certainty. Since the cosine function is bounded, it is also true that $\lim_{t\to\infty} \kappa_{t+1}\cos(\mu_{t+1} - \delta)/t^2 = 0$ with certainty. We have established that $\lim_{t\to\infty} E[\kappa_{t+1}/t^2] = \kappa^2 A^2(\kappa)$, or

$$\lim_{t\to\infty} E\left|\frac{\kappa_{t+1}^2}{\kappa^2 t^2} - A^2(\kappa)\right| = 0 \tag{12}$$

which implies

$$\frac{\kappa_{t+1}^2}{\kappa^2 t^2} \to_{L_1} A^2(\kappa) \Rightarrow \frac{\kappa_{t+1}^2}{\kappa^2 t^2} \to_p A^2(\kappa) \Rightarrow \frac{\kappa_{t+1}^2}{\kappa^2 t^2} \to_d A^2(\kappa).^4 \tag{13}$$

Since the square root function is measurable and continuous over the domain $[0, \infty)$, we also

[4] The converge concepts here are $\to_{L_1}$: convergence in the first moment, $\to_p$: convergence in probability, and $\to_d$: convergence in distribution.

have $\frac{\kappa_{t+1}}{\kappa t} \to_d A(\kappa)$. Since $\frac{\kappa_{t+1}}{\kappa t}$ converges to a constant in distribution, we may conclude that $\lim_{t\to\infty} P\left(\frac{\kappa_{t+1}}{\kappa t} > A(\kappa)\right) = 1 \Rightarrow \lim_{t\to\infty} P\left(\frac{\kappa_{t+1}}{\kappa} > tA(\kappa)\right) = 1$. $\square$

Proposition 2 *In a social belief system with a von Mises distribution, in determining agents' actions, the probability that the weight of public belief (relative to private signals) tending to infinity becomes equal to one as $t \to \infty$.*

Proof. This is a direct indication of Lemma 1.

Remark 1 *Mardia and Jupp (2000) also show that the conditional limiting distribution of μ_{t+1} in this case is von Mises: $\mu_{t+1}|\kappa_{t+1} \sim M(\mu, \kappa_{t+1})$ so that $E[\mu_{t+1}|\kappa_{t+1}] = \mu$ as $t \to \infty$. The expected value of the public belief, given the value of precision of the belief, converges to the true value μ as time goes to infinity, and with κ_{t+1} tending to infinity in the sense of Lemma 1.*

4. The market for information

In this section, we would like to explain why an information provider may have incentives from a public belief with low precision. This incentive could be explained for the case where the information in the hands of the provider is opposing the public belief. It would certainly be easier for the public to absorb the opposing signal when they are not so sure about their existing belief.

One real-life example for this case is a certain type of spam e-mail messages. These messages display tickers of certain stocks and their low price without providing any information about the company, not even its name. Our guess is that such e-mail messages are sent by information providers, not by brokers themselves. Below, we set up a model within our circular framework, and compare our findings with a basic model using a distribution (Gaussian) with linear support.

The linear case

Let us represent the benefits from the signal in terms of the difference it makes in expected squared precision:

$$B_{t+1}(\rho_\epsilon) = bE_t[\rho_{t+1}^2 - \rho^2] = b(\rho_\epsilon^2 + 2\rho_\epsilon\rho_t). \tag{14}$$

where $b > 0$ is the benefit from improvement in unit squared accuracy. Further assume a cost of information retrieval in quadratic form: $C_t(\rho_\epsilon) = p_t\rho_\epsilon^2$, where p_t is the price per unit of squared precision. The condition $MB \geq MC$ will then become:

$$B'_{t+1}(\rho_\epsilon) = 2b(\rho_\epsilon + \rho_t) \geq 2p_t\rho_\epsilon = C'_t(\rho_\epsilon) \Rightarrow \rho_t \geq \left(\frac{p_t}{b} - 1\right)\rho_\epsilon. \tag{15}$$ [5]

For a given precision level ρ_ϵ for sale, the information provider has an incentive to provide a high prior ρ_t to be able to sell s_t with a high price p_t up and above b. This incentive does not explain why we receive spam e-mail promising high profits, with zero precision.

Circular case

Let gains from purchasing information at any time t be represented as the expected gains in squared precision:

$$E_{t_-}[\kappa_{t+1}^2 - \kappa_t^2] = \kappa^2 + 2\kappa\kappa_t E_{t_-}\cos(\mu_t - x_t), \tag{16}$$

For the circular case, we will make a distinction in terms of time t. The beginning of the period when the signal is not received yet will be denoted by t_- and the rest of the period where the value of the signal x_t is known will be denoted by t_+. If this distinction is not relevant, like it is in the linear case, a plain time subscript t will be used for the expectation operator $E[\cdot]$. Now, if we consider the right hand side of (16) times a dollar amount as the benefit of the individual $t+1$ from buying the signal, we may define a benefit function

[5] If we use the gains in terms of smaller posterior variance, then this condition becomes $\left(\frac{p_t}{b} - 1\right)\rho_\epsilon \leq \rho_t < \rho_\epsilon$which further implies $p_t < 2b$ which is still rational. The information provider will surely provide a free prior with a precision less than the precision of the signal for sale.

$B_{t+1}(\kappa)$ as a function of the precision of the signal to be purchased as:

$$B_{t+1}(\kappa) = b[\kappa^2 + 2\kappa\kappa_t E_{t\_} \cos(\mu_t - x_t)]. \tag{17}$$

where $b > 0$ is the benefit from unit increase in expected squared precision of the individual. $B_{t+1}(\kappa)$ is increasing in κ and in the precision of the common belief κ_t if $E_{t\_} \cos(\mu_t - \mu) > 0$, or if information provider's signal x_t is expected not to be too far away from the public belief μ_t.

Now, we will assume a function as the cost function for the retrieval of the signal: $C_t(\kappa) = p_t\kappa^2$, where p_t is the unit price for the square precision purchased at time t. Setting marginal benefit equal to or greater than the marginal cost, we get

$$B'_{t+1}(\kappa) \geq C'_t(\kappa) \Rightarrow \kappa_t E_{t\_} \cos(\mu_t - x_t) \geq \left(\frac{p_t}{b} - 1\right)\kappa. \tag{18}$$

The interpretation of the inequality above complements the finding by Prendergast (1993) that a receiver who is better informed about μ gets better advice, if the provider has an idea about the current belief of the receiver. Our finding here is that a receiver who is better informed (with higher precision κ_t) on μ also *seeks* for more advice in a market for information, if he believes that the signal will be within a certain neighborhood of the mean public belief μ_t satisfying $E_{t\_} \cos(\mu_t - x_t) > 0$.

Suppose the information provider has a fixed precision signal $x_t \sim M(\mu, \kappa)$. In a spam e-mail message promoting a certain financial asset, the information provider communicates two opposing pieces of information: that the stock they want us to buy has a very low price, which reflects the public prior that it is a "bad" stock, and they want us to buy it claiming that it is a "good" stock. These two opposing views could be represented as $E_{t\_} \cos(\mu_t - x_t) = -1$ in our formalism. The receiver of the message will certainly see the two opposing views, and the condition for buying the information will not be satisfied with a p_t up and above b. In the case where $\kappa_t \geq \kappa$, a positive price will not be possible. So the provider will have an incentive to keep the precision of the public prior (κ_t) as low as possible, so that they can maximize the price they will charge for further information which confirms that the stock is noteworthy.

Rewriting the condition (18), we have

$$B'_{t+1}(\kappa) \geq C'_t(\kappa) \Rightarrow b[1 + \frac{\kappa_t}{\kappa} E_{t\_} \cos(\mu_t - x_t)] \geq p_t. \tag{19}$$

Proposition 3 *For a signal with a fixed precision κ for sale, the provider has incentives from the public expectation that the signal value x_t to be close to the mean of the public belief μ_t (as defined by the condition $E_{t\_} \cos(\mu_t - x_t) > 0$), in which case also from the public belief to have high precision.*

If the public expectation is that the signal value x_t will not be close to the mean of the public belief μ_t (as defined by the condition $E_{t\_} \cos(\mu_t - x_t) < 0$), then the provider has incentives from the public belief to have low precision.

The second claim in proposition 3 already explained above. The first claim is also interesting. It states that the provider has incentives to wrap the publicly available information in glossy paper and resell it adding (or perhaps fabricating) new proof in support of the public belief. In the formalism of Prendergast, (1993) the information provider has incentives from being a "yes man". This supporting new signal, when comes with credible (perhaps fabricated) proof, will surely increase the precision of the public belief. The information provider will be putting their credibility on stake if they fabricate new proof supporting the public belief, an event which is not considered in our simplified framework. This possibility of credibility loss, and its effect on the incentives of the information provider calls for further research on the subject within our circular framework.

However, if the information provider succeeds in generating a "self-fulfilling prophecy" through their fabricated signal, they will be gaining credibility, instead of losing credibility.

5. Dynamical learning systems

In this section, we will switch to dynamical systems and derive a circular version of the well known Kalman Filter, and explore the implications of changing the topology of the state-space on social belief systems.

5.1 Learning a dynamical system with dynamical noise

The setting of the social belief system when the underlying subject to be learned is a dynamical system fundamentally differs from the belief systems defined above with only one parameter value to be learned.

The setting is similar to the setting in section 2. The fundamental differences are 1. the public is aware of a model that mimics reality with some uncertainty, and 2. the dynamical variables usually represent the increments of motion rather than the aggregate state variable. The model generates increment $\bar{\theta}_t$ at each period, with a random deviation of $\theta_t - \bar{\theta}_t = \xi_t \sim N(0, \sigma_\xi)$ from the reality. The value of the nature's parameter is therefore $\theta_t = \bar{\theta}_t + \xi_t$. A signal yielding information on the increment then arrives to individual t: $s_t = \theta_t + \epsilon_t$, with ϵ_t is i.i.d. with $N(0, \sigma_\epsilon^2)$. The agents observe each other's actions perfectly, and minimize the squared difference between their actions a_t and the nature's parameter θ_t at each time interval t.

There are two stages of learning in this setting. At the beginning of each period, before the value of the signal is known exactly, individual t adjusts mean y_t and variance σ_t of her belief according to the formulas

$$y_{t_-} = y_{t-1} + \bar{\theta}_{t-1}. \tag{20}$$

The updated belief before the signal is normally distributed. After the signal is received, the belief is again updated,

$$y_t \quad = \quad \alpha_t(s_t + y_{t-1}) + (1-\alpha_t)y_{t\_}. \tag{21}$$

Substituting for $y_{t\_}$, we have

$$y_t = y_{t-1} + \alpha_t(s_t + y_{t-1}) + (1-\alpha_t)\bar{\theta}_t. \tag{22}$$

which implies

$$\triangle y_t = \alpha_t(s_t + y_{t-1}) + (1-\alpha_t)\bar{\theta}_t. \tag{23}$$

So the individuals change the mean of their beliefs relying on the signal for the increment as well as their model. for this case the gain α_t will be time independent: $\alpha = \sigma_\xi^2/(\sigma_\xi^2 + \sigma_\epsilon^2)$. The precision of the change in beliefs will be $\rho_\triangle = \rho_\epsilon + \rho_\xi$, and the variance of change in belief will take the form $\sigma_\triangle^2 = \sigma_\epsilon^2\sigma_\xi^2/(\sigma_\epsilon^2 + \sigma_\xi^2)$. We also know that $\sigma_t^2 = \sigma_{t-1}^2 + \sigma_\triangle^2$ which implies $\sigma_t^2 = \sigma_o^2 + t\sigma_\triangle^2$. So the precision of beliefs $\rho_t = 1/(\sigma_o^2 + t\sigma_\triangle^2)$ converges to zero like $1/t$ as $t \to \infty$.

The dynamical circular case is again fundamentally different from the circular case defined in section 3, however, the setting is similar in many respects. The public is aware of a model that mimics reality with some uncertainty. The model generates $\bar{\delta}_t$ at each period, with a random deviation of $\delta_t - \bar{\delta}_t = \xi_t \sim M(0,\tau)$ from the reality. The nature's parameter has the value $\delta_t = \bar{\delta}_t + \xi_t$. A signal then arrives to individual t: $x_t = \delta_t + \epsilon_t$, with ϵ_t i.i.d. with $M(0,\kappa_\epsilon)$. The agents observe each other's actions perfectly, and minimize the squared difference between their actions a_t and the nature's parameter δ_t at each time interval t.

Similar to the linear case, there are two stages of learning in this setting. At the beginning of each period, after the value of the signal is known exactly, individual t adjusts mean μ_t and variance κ_t of her belief increment according to the formulas

$$R_t^2 = \tau^2 + \kappa_\epsilon^2 + 2\tau\kappa_\epsilon \cos(x_t - \bar{\delta}_t), \tag{24}$$

$$\triangle\mu_{t+1} = \tan^{-1} \frac{\tau \sin \bar{\delta}_t + \kappa_\epsilon \sin x_t}{\tau \cos \bar{\delta}_t + \kappa_\epsilon \cos x_t}. \tag{25}$$

Before the value of the signal x_t is known exactly, the expectation of squared precision of belief takes the form

$$E_{t_-}[R_{t+1}^2] = \tau^2 + \kappa_\epsilon^2 + 2\tau\kappa_\epsilon A(\tau)A(\kappa_\epsilon). \tag{26}$$

The aggregate belief will be approximately distributed as von Mises with precision κ_{t+1} where $A(\kappa_{t+1}) = A(R_t)A(\kappa_t)$. Since $A(R) < 1$ for $0 < R < \infty$ and $A(\cdot)$ is increasing monotonly, κ_{t+1} will converge to zero as $t \to \infty$.

5.2 Optimization in a market for information

When the dynamical social belief system is formulated in terms of increments, the precision of belief on the aggregate variable tends to zero as time goes by. This is due to accumulating uncertainty on the aggregate. Therefore, we assume a consumer of information will be concentrating on their short-run forecasting capability.

Similar to the previous section, we will define a quadratic benefit function representing the benefits from squared expected precision, and a quadratic cost function representing costs associated with increasing the precision of the signal, or precision of the model.

In the Gaussian case, we may represent the benefits from one unit of squared precision as $B(\rho_\triangle) = b(\rho_\epsilon + \rho_\xi)^2$, $b > 0$. The cost functions will be $C_\epsilon(\rho_\epsilon) = c_\epsilon \rho_\epsilon^2$, $c_\epsilon > 0$ and $C_\xi(\rho_\xi) = c_\xi \rho_\xi^2$, $c_\xi > 0$. The optimization problem could then be formulated by the user of the information as

$$\max_{\rho_\epsilon, \rho_\xi} \quad b\rho_\epsilon^2 + b\rho_\xi^2 + 2b\rho_\epsilon\rho_\xi \tag{27}$$

$$s.t. \quad c_\epsilon\rho_\epsilon^2 + c_\xi\rho_\xi^2 \leq M.$$

where $M > 0$ is the budget. The problem above has the interior solution $\rho_\epsilon = \sqrt{\frac{c_\xi}{c_\epsilon}\frac{M}{c_\xi + c_\epsilon}}$, $\rho_\xi = \sqrt{\frac{c_\epsilon}{c_\xi}\frac{M}{c_\xi + c_\epsilon}}$, provided that $b > \frac{c_\epsilon c_\xi}{c_\epsilon + c_\xi}$.

Likewise, in a similar setting for the circular case, we may define an expected benefit function[6] $B(R_t^2) = b[\kappa_\epsilon^2 + \tau^2 + 2\kappa_\epsilon \tau A(\kappa_\epsilon)A(\tau)]$, $b > 0$, and the cost functions as $C_\epsilon(\kappa_\epsilon) = c_\epsilon \kappa_\epsilon^2$, $c_\epsilon > 0$, $C_\tau(\tau) = c_\tau \tau^2$, $c_\tau > 0$.

The optimization problem could then be formulated by the user of the information as

$$\begin{aligned} \max_{\kappa_\epsilon, \tau} \quad & b\kappa_\epsilon^2 + b\tau^2 + 2b\kappa_\epsilon \tau A(\kappa_\epsilon)A(\tau) \\ s.t. \quad & c_\epsilon \kappa_\epsilon^2 + c_\tau \tau^2 \leq M. \end{aligned} \tag{28}$$

which has the interior solution satisfying $\frac{1+\kappa_\epsilon A(\kappa_\epsilon)[1-A^2(\tau)]}{1+\tau A(\tau)[1-A^2(\kappa_\epsilon)]} = \frac{c_\tau}{c_\epsilon}$. Unfortunately, algebraically convenient expressions similar to the linear case are not available for the circular case, and the solutions for precision values are rather cumbersome. However, the case $c_\tau = c_\epsilon = c$ uniquely determines $\tau = \kappa_\epsilon = \sqrt{\frac{M}{2c}}$, $(b > c/2)$ similar to the linear case.

The solutions for the linear and circular cases could be summarized as the higher the unit cost of squared precision, the lower the precision should be, and for equal unit cost of squared precision, equal levels of precision should be employed.

5.3 Chartists vs. fundamentalists

In this section, we will compare a case where the information user entirely relies on her private signal, and the case where establishing research is possible for an overhead cost paid each period, on top of the cost associated with unit squared precision.

The comparison will be made between the maximized profits of the two cases: only signal is hired (Π_1), or both the signal and the model is hired (Π_2) with a research overhead of d paid each period, $C_\xi = d + c_\xi \rho_\xi^2$. For the linear case,

[6] using $E_{t_-}[\cos(x_t - \bar{\delta}_t)] = A(\kappa_\epsilon)A(\tau)$.

$$\begin{aligned}\Pi_1 &= M\frac{b-c_\epsilon}{c_\epsilon}, \\ \Pi_2 &= (M-d)\left(b\frac{c_\epsilon+c_\xi}{c_\epsilon c_\xi}-1\right).\end{aligned} \tag{29}$$

The comparison yields the decision rule: do not invest on research if $\frac{d}{M} > \frac{bc_\epsilon}{bc_\epsilon+(b-c_\epsilon)c_\xi}$, or when $c_\epsilon = c_\xi = c$, do not invest on research if $\frac{d}{M} > \frac{b}{2b-c}$, which turns out to be the decision rule also for the circular case for equal unit costs of squared precision.

6. Conclusion

We have presented a circular framework to investigate the effect of switching from distributions with supports in Euclidean spaces to distributions with supports in non-Euclidean spaces in Bayesian Social Learning Systems. The findings show that there is no difference in terms of the order of convergence (or divergence) between these two specifications, if we switch from linear normal (Gaussian) to circular normal (von Mises) distribution in a basic social learning system.

In a dynamical learning system where both a signal and a stochastic model is available for the forecasting agent, both settings still yield the same conclusion: the sole determinant of the preference between the signal and the model is their cost for unit squared precision. For the equal unit cost, equal precision of signal and model should be purchased.

The findings show that in a simple framework, social learning systems where the agents learn the level of an economic variable, or the phase of the variable is not fundamentally different.

Nevertheless, the circular setting comes with a theoretical gain: a qualitative distinction is embedded into the framework where we can classify the signals continuously between the two extremes of an assertive signal and a negative signal. The incentive behind sending spam e-mails on financial assets bearing no information, and the incentive of the signal provider from being a “yes man” could be explained in this basic framework.

References

Bagchi, P. (1994), "Empirical Bayes estimation in directional data," *Journal of Applied Statistics,* vol.21, pp.317-326.

Barnett, W. A., Geweke, J., Shell, Karl, eds. (1989), *Economic complexity: Chaos, sunspots, bubbles, and nonlinearity.* Proceedings of the Fourth International Symposium in Economic Theory and Econometrics pp. xi, vol. 409, International Symposia in Economic Theory and Econometrics series. Cambridge; New York and Melbourne: Cambridge University Press.

Chamley, C. P. (2004), *Rational Herds,* Cambridge University Press.

Mardia, K., and Jupp, P. E. (2000), *Directional Statistics,* Wiley Series in Probability and Statistics.

Harding, D., Pagan, A. (2006,), "Synchronization of Cycles," *Journal of Econometrics,* v. 132, iss. 1, pp. 59-79.

Prendergast, C. (1993), "A theory of 'yes men,' " *American Economic Review,* vol. 83, pp. 757-770.