In this lecture we define several variants of the *fictitious play* dynamic, and study some of the properties of fictitious play.

Throughout the section we consider a finite N-player game, where each player i has a finite pure action set A_i. We let a_i denote a pure action for player i, and let $s_i \in \Delta(A_i)$ denote a mixed action for player i. We will typically view s_i as a vector in $\mathbb{R}^{A_i}$, with $s_i(a_i)$ equal to the probability that player i places on a_i. We let $\Pi_i(\boldsymbol{a})$ denote the payoff to player i when the composite pure action vector is $\boldsymbol{a}$, and by an abuse of notation also let $\Pi_i(\boldsymbol{s})$ denote the expected payoff to player i when the composite mixed action vector is $\boldsymbol{s}$. We let $BR_i(\boldsymbol{s}_{-i})$ denote the best response mapping of player i; here $\boldsymbol{s}_{-i}$ is the composite mixed action vector of players other than i.

We will typically restrict attention to the case where $N = 2$.

1 Fictitious Play

In this section we introduce three main variants of fictitious play: discrete time fictitious play (DTFP), continuous time fictitious play (CTFP), and stochastic fictitious play (SFP, and the associated perturbed continuous time fictitious play, PCTFP).

1.1 Discrete-Time Fictitious Play

The basic definition of fictitious play was, by most accounts, introduced by Brown [2], although Cournot's best response dynamic is also closely related [3]. Fictitious play refers to a dynamic process where at each stage, players play a (pure) best response to the empirical distribution of their opponent's play. Note that when $N > 2$, this poses a problem: should players respond to the *joint* empirical distribution of their opponents, or the product of the *marginal* empirical distributions of their opponents? Typically, fictitious play considers the product of marginals, although it is worth noting that this is a point of discussion (see [5], Section 2.2).

Let a_i^t denote the action played by player i in time t. The *empirical frequency* of player i's play up to time t is:

$$\gamma_i^t(a_i) = \sum_{\tau=0}^{t-1} \mathcal{I}\{a_i^\tau = a_i\}. \tag{1}$$

Here $\mathcal{I}\{A\}$ denotes the indicator function of A. Thus $\boldsymbol{\gamma}_i^t$ is a A_i-dimensional vector that counts the number of times player i has played each action. The *empirical distribution* of player i's play up to (but not including) time t is:

$$p_i^t(a_i) = \frac{\gamma_i^t(a_i)}{t}. \tag{2}$$

We let $\boldsymbol{p}^t$ denote the distribution on $\prod_i A_i$ given by taken the independent product of the individual distributions p_i^t.

In *discrete-time fictitious play* (DTFP), each player plays an arbitrary action at time 0, and at time t, player i plays a pure best response to the product of the marginal empirical distributions of his opponents; i.e., for all $t > 0$ and i:

$$a_i^t \in BR_i(\boldsymbol{p}_{-i}^t). \tag{3}$$

This is the model that was proposed by Brown and studied by Robinson [2, 9]. Note that, in this model, every player plays a *pure* best response to his opponents.

We make the following remarks:

1. Note that we assume that players move *simultaneously*. Another approach would be to assume that players alternate moves, though almost all the literature takes the simultaneous-move approach. There are not significant qualitative differences between the two models.

2. We have assumed that the initial conditions are specified through an arbitrarily chosen pure action a_i^0. It is often convenient for the purposes of examples to consider instead a situation where players instead begin with initial *beliefs*; in this case, we let p_i^0 denote an initial distribution giving the expectations of other players about player i's play at time 0. At each time $t \geq 0$, we choose $a_i^t \in BR(\boldsymbol{p}_{-i}^t)$, and we let $p_i^t = (\gamma_i^t + p_i^0)/(t+1)$. Effectively, this treats p_i^0 as an arbitrary mixed strategy played by player i at time -1.

3. The update rule that computes p_i^t can be interpreted in a Bayesian sense. Suppose that players other than i initially have a *Dirichlet* prior about player i's play, with parameter vector p_i^t. That is, suppose players believe that the probability that player i will play with the mixed strategy s_i is given by:

$$\mathbb{P}(s_i) = \frac{1}{\prod_{a_i \in A_i} \Gamma(p_i^t(a_i))} \prod_{a_i \in A_i} s_i(a_i)^{p_i^t(a_i)-1}.$$

   Here Γ is the gamma function. It is straightforward to check that in this model, the expected value of $s_i(a_i)$ is exactly $p_i^t(a_i)$. Further, after player i plays at time t, suppose that action a_i^t is realized. It is also straightforward to show that the posterior (or conditional) distribution of s_i, given that player i played a_i^t, is a Dirichlet distribution with parameter vector p_i^{t+1}. We'll revisit this Bayesian approach when we study learning.

4. We assume here that players put equal weight on every play in the past. An alternative approach is to consider weightings that prioritize recent play. One extreme, where $p_i^t(a_i^{t-1}) = 1$, and $p_i^t(a_i) = 0$ for all other a_i, is called the *best response dynamic*. (Note that this terminology is not canonical; some authors even refer to fictitious play as the best response dynamic.) The best response dynamic is the same one first considered by Cournot. We will revisit its performance when we study supermodular games.

1.2 Continuous-Time Fictitious Play

We now consider a continuous-time version of the process in the previous section. In *continuous-time fictitious play* (CTFP), players update the weights they place on the actions available to them, in the *direction* of a best response to their opponents' past actions. Formally, we have:

$$\frac{dp_i^t}{dt} \in BR(\boldsymbol{p}_{-i}^t) - p_i^t. \tag{4}$$

(Note the left hand side is an A_i-dimensional vector.) To interpret this equation, it is helpful to think of p_i^t as the current empirical distribution of the opposing players' play. Player i "adjusts" his play so that his empirical distribution has a drift towards an element of the best response to $\boldsymbol{p}_{-i}^t$.

Processes such as (4) are called *differential inclusions*. It is straightforward but technically involved to establish that such inclusions have a solution, under reasonable conditions. We refer to a process $\{\boldsymbol{p}^t\}$ that solves the inclusion (4) as a CTFP process. However, we emphasize that because selections from the best response mapping are typically highly discontinuous, there may exist *many* solutions to the CTFP differential inclusion. (See [7] for a more detailed discussion of this issue.)

It is not difficult to establish that CTFP is, in an appropriate sense, a continuous time version of DTFP. Suppose we are given a CTFP process $\{\boldsymbol{p}^t\}$ defined on $(-\infty, \infty)$, and define $s_i^t = p_i^t + dp_i^t/dt$. It follows from the definition of CTFP that:

$$s_i^t \in BR(\boldsymbol{p}_{-i}^t). \tag{5}$$

Further, we can establish that:

$$p_i^t = \int_{-\infty}^{t} e^{-(t-\tau)} s_i^\tau \, d\tau. \tag{6}$$

(Simply differentiate the preceding expression and substitute the definition of s_i^t.) We can thus view s_i^t as the mixed action profile chosen by player i at time t, and p_i^t as the exponentially weighted empirical distribution of player i's past history of play. In fact, this relationship holds in reverse: given processes $\boldsymbol{s}^t$ and $\boldsymbol{p}^t$ that satisfy (5)-(6), the process $\boldsymbol{p}^t$ is a CTFP process.

We conclude with one more transformation. If we change time units, letting $\hat{p}_i^t = p_i^{\log t}$ and $\hat{s}_i^t = s_i^{\log t}$, then it follows that for $t \geq 0$ and all i we have:

$$\hat{s}_i^t \in BR(\hat{\boldsymbol{p}}_{-i}^t),$$

and

$$\hat{p}_i^t = \frac{1}{t} \int_0^t \hat{s}_i^{t\hat{a}u} \, dt\hat{a}u.$$

(We use hats here to emphasize the change of variables required to establish this identity.) Thus, with an additional change of time units, CTFP is exactly a continuous time generalization of the DTFP process. Note, however, that this analog tracks the past *mixed* actions of player i, not the realized *pure* actions; nevertheless, as Harris shows [7], the preceding relation implies that asymptotic behavior of CTFP can be used to infer asymptotic behavior of DTFP.

1.3 Stochastic Fictitious Play

DTFP is somewhat strange in its original form, because players only play pure actions at each time step; this is true even if players actually have in mind a mixed best response. This means that actual play can never converge to a mixed Nash equilibrium. Even if we track players' mixed actions, the best response mapping may still be multiple valued, raising significant hurdles in the analysis of fictitious play.

One approach to "smoothing" fictitious play is to introduce randomness into the model. Formally, Fudenberg and Kreps took an approach similar to Harsanyi's purification of Nash equilibrium to define stochastic fictitious play [4]. Harsanyi assumed that in a standard simultaneous-move game, each player privately observed a "shock" to their payoffs—a small noise term that shifted the payoffs slightly. Since the shock to player i's payoffs is only observed by player i, this makes the original complete information game into a game of incomplete information. Harsanyi's purification theorem shows that (under some assumptions) any mixed strategy equilibrium of the original complete information game is a limit of pure strategy Bayesian equilibria of the perturbed incomplete information games, as the noise term approaches zero. (Details can be found in any standard game theory text, e.g., [6].)

1.3.1 The Choice Probability Function and SFP

We can consider the same approach in the fictitious play model. We follow the development of Hofbauer and Sandholm [8]. Suppose that at each time period, the players privately observe a random shock to their payoffs; in particular, we let $\Pi_i(\boldsymbol{a}) + \varepsilon N_i$ denote the payoff to player i when the composite action is $\boldsymbol{a}$ and the random shock N_i is realized. (In general, the random shock could depend on the action chosen by player i, or even on the composite vector of actions $\boldsymbol{a}$; all that is important is that only player i observes the functional form of the shock to his own payoff.)

In *stochastic fictitious play* (SFP), at time t, player i chooses an action that maximizes his perturbed payoff, assuming that each opponent plays according to his (marginal) empirical distribution of play up to time t. Note that the distribution over N_i gives rise to a distribution over the possible actions that will be played by player i at time t; in particular, we define the distribution $C_i(\boldsymbol{s}_{-i})$ according to:

$$C_i(\boldsymbol{s}_{-i})(a_i) = \mathbb{P}\left(\arg\max_{a_i' \in A_i}[\Pi_i(a_i', \boldsymbol{s}_{-i}) + \varepsilon N_i] = a_i\right).$$

Note that for this to be well defined, we require that the noise has strictly positive density everywhere in $\mathbb{R}$.

The distribution $C_i(\boldsymbol{s}_{-i})$ is called the *choice probability function* of player i; it gives the probability that player i will choose any one of actions, given the composite (mixed) action vector of his opponents. For this reason C_i is also called the *perturbed best response function* of player i; under this interpretation, we view $C_i(\boldsymbol{s}_{-i})$ as the perturbed-payoff-maximizing mixed action of player i, when his opponents play $\boldsymbol{s}_{-i}$.

In comparing with (3), we see that in stochastic fictitious play, the initial action a_i^0 is arbitrary, and the action a_i^t is randomly drawn according to the distribution specified by the perturbed best response $C_i(\boldsymbol{p}_{-i}^t)$.

1.3.2 Perturbed CTFP

Using the choice probability function, we can compute the distribution governing the evolution of SFP. Define the empirical frequency and distribution of player i's play as in the discrete-time equations, (1) and (2), respectively. Observe that:

$$p_i^{t+1} = \frac{\gamma_i^{t+1}}{t+1} = \frac{1}{t+1}\left(tp_i^t + \mathcal{I}_{a_i^{t+1}}\right),$$

where by an abuse of notation we use $\mathcal{I}_{a_i}$ to denote the indicator function of the action a_i. We note that conditional on $\boldsymbol{p}^t$, the action a_i^{t+1} is drawn according to the choice probability distribution $C_i(\boldsymbol{p}_{-i}^t)$. It follows that the expected change in player i's empirical distribution is:

$$\mathbb{E}[p_i^{t+1} - p_i^t|\boldsymbol{p}^t] = \frac{1}{t+1}\left(\mathbb{E}[\mathcal{I}_{a_i^{t+1}}|\boldsymbol{p}^t] - p_i^t\right) = \frac{1}{t+1}\left(C_i(\boldsymbol{p}_{-i}^t) - p_i^t\right).$$

By analogy with our discussion of CTFP, the previous calculation suggests a natural analog, which we call *perturbed continuous time fictitious play* (PCTFP):

$$\frac{dp_i^t}{dt} = C_i(\boldsymbol{p}_{-i}^t) - p_i^t. \tag{7}$$

Note the similarity to the CTFP process (4): the best response function has been replaced by the choice probability function C_i. Because the choice probability function is uniquely defined, the PCTFP dynamics are a differential *equation*, rather than a differential *inclusion*.

1.3.3 Reinterpreting the Choice Probability Function

We conclude with a reinterpretation of the choice probability function that helps both intuition and analysis. It turns out that a convenient simplification is possible from the random utility model described above. Given *any* (continuously differentiable) choice probability function C_i, there exists a convex function V_i on $\Delta(A_i)$ such that:

$$C_i(\boldsymbol{s}_{-i}) = \arg\max_{s_i\in\Delta(A_i)} \left(\Pi_i(s_i, \boldsymbol{s}_{-i}) - V_i(s_i)\right).$$

(Note that the calculation on the right hand side is the standard Legendre transform of convex analysis.) One common example is the *logit choice probability function*: if $V_i(s_i) = -\varepsilon\sum_{a_i} s_i(a_i)\log s_i(a_i)$ (negative entropy), then the resulting C_i is:

$$C_i(\boldsymbol{s}_{-i})(a_i) = \frac{\exp\left(\Pi_i(a_i, \boldsymbol{s}_{-i})/\varepsilon\right)}{\sum_{a_i'\in A_i}\exp\left(\Pi_i(a_i', \boldsymbol{s}_{-i})/\varepsilon\right)}.$$

This choice function comes up in a variety of fields: information theory, statistics, physics, large deviations, etc.

The preceding transformation is useful, because it allows us to represent the PCTFP dynamics using a *deterministic* perturbation V_i. Further, it is straightforward to interpret the effect of V_i analytically: while maximization without the perturbation term yields a discontinuous best response mapping, the perturbation "softens" the maximum; in particular, it guarantees that the perturbed best response C_i is always unique, regardless of opponents' strategies. It is straightforward to check that as the coefficient ε of N_i approaches zero, the perturbation V_i approaches zero as well. In this case the function C_i more and more closely resembles the true best response mapping; for example, it is straightforward to establish that as $\varepsilon \to 0$, the logit choice function chooses only those actions that maximize the unperturbed payoff.

2 Nash Equilibrium

A basic insight behind the definition of fictitious play is that if it "converges", then it must converge to a Nash equilibrium of the game. In this section we formalize this insight.

2.1 DTFP

We say that DTFP *converges* if the empirical distributions p_i^t converge for every player i, as $t \to \infty$. If a fixed pure action profile $\boldsymbol{a}^t$ is played at every time $t \geq t_0$, then we call $\boldsymbol{a}^t$ a *steady state* of DTFP.

It is immediate that if $\boldsymbol{a}^t$ is a steady state of DTFP, then it must be a Nash equilibrium. This is because eventually p_i^t must also converge to a point mass on the action a_i^t; and thus if a_i^t is not a best response to $\boldsymbol{a}_{-i}^t$, eventually player i will choose a different action.

It is also similarly obvious that if $\boldsymbol{a}$ is a strict Nash equilibrium—i.e., if a_i is a strict best response to $\boldsymbol{a}_{-i}$—then once $\boldsymbol{a}$ is played at time t, it is played at all subsequent times thereafter. To see this, note that if player i plays a_i at time t, then $a_i \in BR_i(\boldsymbol{p}_{-i}^t)$; in particular, for all a_i':

$$\Pi(a_i, \boldsymbol{p}_{-i}^t) \geq \Pi(a_i', \boldsymbol{p}_{-i}^t).$$

But all players other than i play $\boldsymbol{a}_{-i}$ at time t, and by assumption we have for all a_i':

$$\Pi(a_i, \boldsymbol{a}_{-i}) > \Pi(a_i', \boldsymbol{a}_{-i}).$$

Since p_j^{t+1} is a convex combination of p_j^t and a point mass on a_j^t, we conclude that for all a_i':

$$\Pi(a_i, \boldsymbol{p}_{-i}^{t+1}) > \Pi(a_i', \boldsymbol{p}_{-i}^{t+1}).$$

Thus player i will play a_i at all times $t' \geq t$.

Finally, suppose that DTFP converges, i.e., $p_i^t \to p_i$ as $t \to \infty$ for all i. The key insight is that *the limiting product of empirical distributions* $\boldsymbol{p}$ *is a Nash equilibrium*. Suppose not; in particular, suppose for some player i and pure action a_i we have:

$$\Pi_i(a_i, \boldsymbol{p}_{-i}) > \Pi_i(\boldsymbol{p}).$$

Then for all sufficiently large t, say for $t \geq t_0$, we have:

$$\Pi_i(a_i, \boldsymbol{p}_{-i}^t) > \Pi_i(p_i, \boldsymbol{p}_{-i}^t).$$

But then for all $t \geq t_0$, player i would never play any of the actions on which he places positive weight in p_i. This contradicts the assumption that the empirical distribution of player i converges to p_i.

It is the preceding simple argument that justifies fictitious play: it is the most natural update rule in which a player plays a best response to his opponent's behavior (or perhaps second best, after best response dynamics). Thus, asymptotically, if DTFP converges, all players are playing best responses to each other – a Nash equilibrium.

2.2 CTFP

We define convergence for CTFP in the obvious way: we say that a CTFP process $\boldsymbol{p}^t$ *converges* if p_i^t converges for every player i, as $t \to \infty$. Since this also implies that $dp_i^t/dt \to 0$ as $t \to \infty$, we conclude that the limiting empirical distribution $\boldsymbol{p}$ must satisfy $p_i \in BR_i(\boldsymbol{p}_{-i})$; i.e., it is a Nash equilibrium, matching the discrete-time result.

2.3 SFP and PCTFP

Convergence for SFP is slightly more subtle, since it defines a stochastic process rather than a deterministic dynamical system. For simplicity, we only define convergence here for games that admit a unique Nash equilibrium; in general, convergence for stochastic play must be defined by considering the limit points of the trajectory, and we refer the reader to [8] as well as [5], Chapter 4, for the details.

For our purposes, we will say that the SFP dynamics *converge* to $\boldsymbol{p}$ if $\mathbb{P}(\boldsymbol{p}^t \to \boldsymbol{p}) = 1$ (i.e., the empirical distributions almost surely converge to $\boldsymbol{p}$). Suppose that the perturbed game—i.e., the game where each player maximizes $\Pi(\boldsymbol{s}) + V_i(s_i)$—has a unique (possibly mixed) Nash equilibrium $\boldsymbol{s}$. (We will later see conditions under which we can guarantee this.) Then by reasoning analogous to the discrete-time setting, it follows that if SFP converges to $\boldsymbol{p}$, then $\boldsymbol{p}$ must be the unique Nash equilibrium $\boldsymbol{s}$.

By reasoning analogous to the continuous-time setting, we can also see that if PCTFP converges (i.e., if the resulting process $\boldsymbol{p}^t$ converges to a limit $\boldsymbol{p}$), then the limiting empirical distributions $\boldsymbol{p}$ must be the unique Nash equilibrium of the perturbed game. As the size of the perturbation ε approaches zero, we can establish that the limiting empirical distributions approach a Nash equilibrium of the original game (see Proposition 3.1 in [8]).

We conclude with a comment on the relationship between convergence of SFP and convergence of PCTFP. Note that in SFP, we can write the update of empirical distributions as follows:

$$p_i^{t+1} - p_i^t = \frac{C_i(\boldsymbol{p}_{-i}^t) - p_i^t}{t+1} + \frac{\xi_i^t}{t+1},$$

where $\xi_i^t = \mathcal{I}_{a_i^t} - C_i(\boldsymbol{p}_{-i}^t)$ is a zero mean random variable. Now as $t \to \infty$, the last term on the right approaches zero; further, the cumulative variance of the noise terms is bounded, since $\sum_t 1/t^2 < \infty$. Informally, this suggests that the limiting behavior of the *stochastic* process $\boldsymbol{p}^t$ can be inferred from the limiting behavior of the following *deterministic* process:

$$\hat{p}_i^{t+1} - \hat{p}_i^t = \frac{C_i(\hat{\boldsymbol{p}}_{-i}^t) - \hat{p}_i^t}{t+1}.$$

In turn the limiting behavior of the preceding deterministic process is characterized by the limiting behavior of the PCTFP process. Thus we conjecture that if the PCTFP process converges to a unique limit $\boldsymbol{p}^*$ regardless of initial condition, then the SFP dynamics should converge almost surely to the same limit.

Indeed, this is the approach of the theory of *stochastic approximations*, a tool that has found great use in studying stochastic fictitious play. Fudenberg and Kreps apply this approach to study 2×2 matrix games with a unique Nash equilibrium [4]. Of course, the story is significantly more complicated in general, and in particular when the PCTFP process is not globally asymptotically stable. For further details, we refer the reader to [1, 8] as well as [5], chapter 4.

2.4 A Warning on Convergence Notions

We close with a warning on the notions of convergence used in the analysis of fictitious play. Typically, the desired convergence is in the empirical distributions of the players. However, it is frequently the case that results are proved instead for the payoffs of the players; in particular, a common statement is that fictitious play "converges" if for all i:

$$\lim_{t\to\infty} \max_{s_i \in \Delta(A_i)} \Pi_i(s_i, \boldsymbol{p}_{-i}^t) - \Pi_i(\boldsymbol{p}^t) = 0.$$

This is convergence of payoffs: i.e., the payoff at the empirical distribution for player i approaches the best possible payoff player i could achieve against stationary opponents playing $\boldsymbol{p}_{-i}^t$. Payoff convergence is a useful notion if the empirical distributions may oscillate, even when payoffs have stabilized.

References

[1] M. Benaïm and M. Hirsch. Mixed equilibria and dynamical systems arising from fictitious play in perturbed games. *Games and Economic Behavior*, 29:36–72, 1999.

[2] G. Brown. Iterative solution of games by fictitious play. In T. Koopmans, editor, *Activity Analysis of Production and Allocation*. John Wiley and Sons, New York, New York, 1951.

[3] A. A. Cournot. *Recherches sur les Principes Mathématiques de la Théorie des Richesses*. Macmillan and Co., New York, New York, 1897. Translated by N.T. Bacon; originally published in 1838.

[4] D. Fudenberg and D. M. Kreps. Learning mixed equilibria. *Games and Economic Behavior*, 5:320–367, 1993.

[5] D. Fudenberg and D. K. Levine. *The Theory of Learning in Games*. MIT Press, Cambridge, Massachusetts, 1998.

[6] D. Fudenberg and J. Tirole. *Game Theory*. MIT Press, Cambridge, Massachusetts, 1991.

[7] C. Harris. On the rate of convergence of continuous-time fictitious play. *Games and Economic Behavior*, 22:238–259, 1998.

[8] J. Hofbauer and W. Sandholm. On the global convergence of stochastic fictitious play. *Econometrica*, 70:2265–2294, 2002.

[9] J. Robinson. An iterative method of solving a game. *Annals of Mathematics*, 54:296–301, 1951.