Statistical Inference in Games *

Yuval Salant                Josh Cherry
Northwestern University      Amazon.com

February 20, 2019

Abstract

We consider statistical inference in games. Each player obtains a small random sample of other players’ actions, uses statistical inference to estimate their actions, and chooses an optimal action based on the estimate. In a Sampling Equilibrium with Statistical Inference (SESI), the sample is drawn from the distribution of players’ actions based on this process. We characterize the set of SESIs in large two-action games, and compare their predictions to those of Nash Equilibrium, and for different sample sizes and statistical inference procedures. An application to search and matching markets demonstrates that statistical inference from small samples leads to significantly larger unemployment — and significantly larger employment fluctuations in response to exogenous shocks — than in Nash Equilibrium.

*We thank Eduardo Azevedo, Joe Harrington, Emir Kamenica, Shengwu Li, Sanket Patil, Neal Roese, Ariel Rubinstein, Ron Siegel, Ran Spiegler, and seminar participants in Bar-Ilan University, Ben-Gurion University, Brown University, Hebrew University, Northwestern University, Penn-State University, Tel-Aviv University, University of Pennsylvania, University of Toronto, University of Western Ontario, the Advances in Behavioral Game Theory conference at University College London, the Applied Behavioral Theory conference at the University of Zurich, the BRIC conference, and the Recent Developments in Behavioral Economics and Mechanism Design conference in the University of Glasgow for helpful discussions and comments.
1 Introduction

Real-life decision making often involves deciding whether to take a particular action or not in situations in which the action’s value depends on how many other people take the action. For example, deciding whether to drive to work may depend on road congestion; when contemplating whether to search for a new job, market thickness is a relevant factor; and deciding whether to purchase a good may depend on other people’s demand when consumption generates externalities. In such settings, an individual has to determine how many other people take the action in order to figure out the action’s value.

Individuals may act as statisticians when estimating how many people take the action. An individual may first obtain some data on other people’s actions. For example, the individual may ask a few co-workers about whether they drive to work. The individual may then use some form of statistical inference to estimate how many people take the action. For example, the individual may combine the data collected with some prior perception he has about road congestion, or perhaps calculate the most likely road congestion parameter to have generated the data. Based on his estimate, the individual can then decide whether to take the action or not.

This paper considers such statistical decision making in games. Each player obtains small sample of other players’ actions. To describe how players make inferences from the sample, we introduce the notion of an inference procedure, which is the analogue of an estimator in the statistics literature. An inference procedure assigns to every possible sample a player may obtain an estimate, which is a distribution over possible proportions of players taking the action. Estimates are assumed to relate to one another monotonically: fixing the sample size, a player puts a larger weight on more players taking the action as the number of observations in which people take the action increases. Examples of inference procedures include Bayesian updating from a non-degenerate prior, maximum likelihood estimation, and Beta estimation, among others. An inference procedure can also be used to describe various heuristics in information processing such as the law of small numbers (Tversky and Kahneman (1971)), according to which decision makers tend to underestimate sample variability. Based on the estimate, each player chooses an action that maximizes his payoff.

A sampling equilibrium with statistical inference (SESI) incorporates this statistical decision making procedure into Osborne and Rubinstein (2003)’s sampling equilibrium. A SESI is a distribution of actions with the property that sampling from this distribution and using statistical
inference to arrive at an optimal action results in the same distribution of actions. A SESI depends on individuals’ sample sizes and inference procedures.

In a SESI, players have access to the equilibrium distribution of actions. Osborne and Rubinstein (1998), who were the first to make this assumption, interpret it as reflecting a steady state of a dynamic process in which new players sample the actions of past players. Sethi (2000) formalizes this interpretation in a dynamic environment and studies its implications for equilibrium selection. A variation of this interpretation is that of noisy retrieval of information from memory. A player may have been involved in the same or a similar interaction in the past and may have encoded in memory the distribution of actions back then. However, because of memory decay and forgetting, the player can only retrieve from memory imperfect signals about his past experience. In this interpretation, Nash equilibrium reflects an ability to perfectly retrieve information from memory in contrast to the imperfect retrieval in a SESI.

Section 3 studies SESIs in large games in which the action players consider taking has an idiosyncratic benefit and a cost that is increasing and convex in the number of players taking the action. We establish three main results, which then serve as building blocks for studying applications in later sections. First, for any given inference procedure and sample size, there is a unique SESI. Second, the proportion of players who take the action in this SESI, which we call the SESI proportion, is smaller than the Nash equilibrium proportion for any sample size and for any unbiased inference procedure, i.e., an inference procedure with the property that the expected value of the estimate is the sample mean. The SESI proportion could be as small as half of the NE proportion for natural cost functions. Third, under additional assumptions, the SESI proportion increases in the sample size.

Section 4 studies an application to competitive markets. A unit mass of producers face a known demand function for their product, and each has to decide whether to produce a unit of the product at an idiosyncratic cost. To do so, each producer has to estimate the market price, which depends on the production decisions of other producers. Each producer thus obtains information on the production decisions of a few other producers, and uses statistical inference to estimate the market supply and the market price.

---

1 A classic contribution to the study of memory decay and forgetting is Ebbinghaus (1885) (See Ebbinghaus (2003) for the English version.) A more recent contribution is Schacter (1999, 2002).

2 A possible difficulty with the information retrieval interpretation is that we assume the player’s samples are independent of one another. Vul and Pashler (2008) provide partial support for this independence assumption.
Because there is a unit mass of producers, their data about others’ production decision — when aggregated over all producers — reflects the market supply accurately. But because each producer only obtains a small sample, some producers will underestimate and some will overestimate the market supply. This is in contrast to a standard competitive equilibrium in which each producer estimates the market supply accurately. We establish that there is a unique SESI in this environment, and that the market supply in this SESI depends on the curvature of the inverse demand function. In particular, when the inverse market demand is convex, there is over-production and a lower market price in the unique SESI than in the standard competitive equilibrium.

Section 5 considers an application to monopoly pricing in a market in which consumers have a preference for uniqueness, i.e., the more consumers who buy the product the less valuable it is. A leading example is status goods. We ask how the monopolist’s optimal profit changes as consumers obtain more information about other consumers’ purchase decisions. We establish that the monopolist’s profit decreases in the amount of information consumers obtain, thus suggesting that the monopolist has an incentive to reduce consumers’ access to information about other consumers’ purchase decisions.

Section 6 considers an application to search and matching markets. Workers in the market have to decide whether to engage in costly search for a job, and firms have to decide whether to engage in costly search for a worker. A Cobb-Douglas matching function determines the likelihood of a worker-firm match as a function of the matching friction and the market thickness on the worker side and the firm side. As pointed out by Petrongolo and Pissarides (2001), “the stylized fact that emerges from the empirical literature is that there is a stable aggregate matching function of a few variables that satisfies the Cobb-Douglas restrictions with constant returns to scale in vacancies and unemployment.”

We depart from the rational expectation modeling assumption by assuming that workers (and similarly firms) do not know the thickness on the other side of the market and use statistical inference to estimate it. A SESI in this environment is a steady state in which workers obtain data from the distribution of firms’ actions based on the firms’ statistical inference, and firms obtain data from the distribution of workers’ actions based on workers’ statistical inference.

Our main result is about market thickness and employment in SESIs. We show that there is a unique SESI with positive employment. In this SESI, market thickness and employment are smaller than in the unique rational expectations equilibrium with positive employment. The gap
in market thickness and employment can be proportional to the matching friction itself. This implies that statistical inference amplifies the effect of exogenous shocks on employment relative to the predictions of the standard model with rational expectations.

Section 7 concludes with a discussion of heterogeneity in players’ sample sizes and inference procedures. Heterogeneity along these dimensions may arise due to a variety of reasons such as differences in players’ ability to obtain or process information, different time constraints players are facing, or different preferences over statistical inference methods. We establish that our existence and uniqueness results as well as the comparison to Nash equilibrium extend to environments with heterogeneous players. We also identify simple conditions on the profile of players’ sample sizes and inference procedures under which our comparative statics results continue to hold.

1.1 Related literature

In Nash equilibirum, players are assumed to have the correct beliefs about the equilibrium distribution of actions. Dissatisfaction with this assumption has led to the development of alternative solution concepts. Prominent examples include self-confirming equilibrium (Fudenberg and Levine (1993)), sampling equilibrium (Osborne and Rubinstein (1998, 2003)), cursed equilibrium (Eyster and Rabin (2005)), analogy-based equilibrium (Jehiel (2005)), behavioral equilibrium (Esponda (2008)), and Berk–Nash equilibrium (Esponda and Pouzo (2016)). Of this literature, Osborne and Rubinstein (1998, 2003) are the most relevant to the current paper.

Osborne and Rubinstein (1998) consider games with players who sample the payoff of each of their actions and choose the action with the largest payoff. They introduce the notion of $S(k)$-equilibrium whereby the distribution from which players sample is identical to the distribution of payoffs based on this process, and analyze the predictions of $S(1)$-equilibria in two-player games. Osborne and Rubinstein (2003) consider the notion of a sampling equilibrium in which players sample the actions of other players instead of their own payoffs and best-respond to sample averages. They analyze the sampling equilibria of a voting model with two or three samples. Sethi (2000) studies the dynamic stability of $S(1)$-equilibria, and Mantilla, Sethi, and Cárdenas (2018) analyze the efficiency and stability of $S(1)$-equilibria in public goods games. Spiegler (2006a,b) studies firms’ optimal behavior when they face a population of consumers who sample their prices or another payoff-relevant parameter once.
We make three contributions to this literature. First, we consider players who use statistical inference to form beliefs about the equilibrium distribution of actions. Statistical inference is captured by the notion of an inference procedure, which is flexible enough to model a rich class of possible statistical inferences from data. A related contribution is Liang (2018) in which players with different learning rules observe the same long sequence of data on a payoff-relevant parameter they wish to estimate. Liang (2018) studies whether the model predictions converge to Nash equilibrium for large data sequences. Other related contributions are Al-Najjar (2009) who studies statistical decision making in a single-person environment, and Spiegler (2018) who studies two-player simultaneous-move games in which players with access to infinite data on correlations between payoff-relevant variables use the maximum entropy criterion to form beliefs about the opponent’s action.

Our second contribution is to provide a comprehensive characterization of SESIs in large two-action games. Our ability to characterize SESIs relies on the connection between SESIs and Bernstein polynomials that we discuss in detail in Section 3.1. Nöldeke and Peña (2016) and Peña, Lehmann, and Nöldeke (2014) are earlier contributions that use Bernstein polynomials in other game-theoretic settings to study Nash equilibria of symmetric voter participation games and evolutionary dynamics in two-action $N$-player normal form games.

Our third contribution is to develop applications to competitive markets, monopoly pricing, and search and matching markets.

2 Model

This section presents the model. After describing the model primitives in Section 2.1, Section 2.2 discusses players’ statistical decision making. Section 2.3 discusses the data generating process and the solution concept. Section 2.4 provides an example.

2.1 Primitives

There is a unit mass of players, and each of them has to decide whether to take action $A$ or action $B$. The utility from action $B$ is 0. The utility from action $A$ is $u(\theta, \alpha) = \theta - f(\alpha)$, where $\theta$ is a player’s idiosyncratic benefit from $A$, and $f(\alpha)$ is the cost incurred by a player taking the action $A$ if a proportion $\alpha$ of players take the action $A$. The benefit $\theta$ is distributed uniformly
on $[0, 1]$, and the function $f$ is continuous, convex and increasing with $0 \leq f(0)$ (i.e., $f$ is a cost) and $f(1) \leq 1$ (i.e., the cost is weakly smaller than the maximal benefit of action $A$.)

### 2.2 Statistical decision making

In order to decide which action to take, each player has to estimate the proportion $\alpha$ of players taking the action $A$. To do so, the player obtains $k$ observations of other players’ actions, which are $k$ independent draws from a Bernoulli distribution with probability of success $\alpha$. Observing a success is interpreted as observing a player who takes the action $A$, and a failure as observing a player who takes the action $B$. The resulting sample is denoted by $(k, z)$ where the integer $k \geq 1$ is the sample size and $z \in [0, 1]$ is the sample mean, i.e., the proportion of successes in the sample.³

An inference procedure, which is the analogue of an estimator in the statistics literature, describes how the player makes inferences from the sample. It requires that the player puts more weight on more players taking the action $A$ as the sample mean increases.

**Definition (Inference Procedure).** An inference procedure $G = \{G_{k,z}(\cdot)\}$ assigns a cumulative distribution function $G_{k,z}(\cdot)$, called an estimate, to every sample $(k, z)$ such that the estimate $G_{k,z}(\cdot)$ strictly first-order stochastically dominates the estimate $G_{k,z}(\cdot)$ whenever $\hat{z} > z$.

Following are a few examples of inference procedures.

**Example 2.1** (Bayesian Updating). A player has a non-degenerate prior over the parameter $\alpha$, and he uses Bayes rule to update this prior based on the sample. By Proposition 1 in Milgrom (1981), the family of posteriors is an inference procedure.

**Example 2.2** (Maximum Likelihood Estimation (MLE)). A player uses the maximum likelihood method to estimate the most likely parameter $\alpha$ to generate the sample, i.e., the player solves for the $\alpha$ that maximizes $\alpha^k z (1 - \alpha)^{k(1 - z)}$. It is easy to verify that this $\alpha$ is equal to $z$. The player treats this number as the proportion of players taking the action $A$. Thus, the inference procedure is

$$G_{k,z}(\alpha) = \begin{cases} 
0 & \alpha < z \\
1 & \alpha \geq z
\end{cases}$$

³The proportion $z$ can take any value in $[0, 1]$ even though players only observe proportions of the form of $j/k$ for $j = 0, 1, ..., k$. This is useful for comparative statics.
Example 2.3 (Beta Estimation). A player has “complete ignorance” about the proportion $\alpha$ taking the action $A$. Such ignorance is often captured in the statistics literature by “Haldane’s prior” (Haldane, 1932; Zhu and Lu, 2004), which is the limit of the Beta($\epsilon, \epsilon$) distributions as $\epsilon \to 0$.

When the player’s sample includes only failures (i.e., $z = 0$) or successes (i.e., $z = 1$), the player concentrates his estimate on the sample mean similarly to MLE. When the sample includes both successes and failures (i.e., $0 < z < 1$), the player’s estimate is the Beta($zk, (1-z)k$) distribution. The mean of this estimate is $z$ and its variance is $\frac{z(1-z)}{k+1}$.

To understand why a player with “complete ignorance” who observes the sample $(k, z)$ may arrive at the estimate Beta($zk, (1-z)k$), recall that the Beta distribution is a conjugate prior for Binomial distributions. That is, given a Beta($a,b$) prior on $\alpha$, the posterior following the realization of a Binomial random variable with parameters $k$ (sample size) and $\alpha$ (probability of success) is a Beta($a+s, b+k-s$) distribution, where $s$ is the number of successes in the sample. That is, the first parameter of the Beta is incremented by the number of successes, and the second by the number of failures. For example, the uniform distribution is a Beta(1,1) distribution, and so a player with a uniform prior who observes the sample $(k, z)$ and uses Bayesian updating would have the posterior Beta($1 + zk, 1 + (1-z)k$).

The sum $a + b + k$ of the two Beta parameters is often interpreted as the number of pseudo-observations, a way of measuring the weight placed on the prior ($a + b$) relative to that put on the sample $k$. Thus, a uniform prior corresponds to putting a weight of 2 on the prior and the rest on the sample, whereas in the current example, players are ignorant in the sense that they put no weight on a prior and base their estimate solely on the sample.

Note that Beta estimation is not included in Example 2.1 because there does not exist a proper prior that, together with Bayesian updating based on the sample $(k, z)$, generates a Beta($zk, (1-z)k$) posterior.

Example 2.4 (Truncated Normal). A player believes that the proportion taking action $A$ is distributed according to a normal distribution truncated symmetrically around the mean. He equates the mean of the distribution with $z$ and the variance with $\frac{z(1-z)}{k}$.

After obtaining the sample $(k, z)$ and deriving the estimate $G_{k,z}(\cdot)$, a player best-responds
to his estimate. That is, he takes the action $A$ if and only if

$$\theta \geq F_{k,z}$$

where $F_{k,z}$ is the expected value of the cost $f$ under $G_{k,z}(\cdot)$, i.e.,

$$F_{k,z} = \int_0^1 f(\alpha) dG_{k,z}(\alpha).$$

### 2.3 Equilibrium

Players obtain their sample from the distribution of actions based on players’ statistical decision making. Put differently, players obtain data from a source that is representative of the parameter they wish to estimate. A sampling equilibrium with statistical inference describes the resulting solution concept.

**Definition** (Sampling Equilibrium with Statistical Inference). A **sampling equilibrium with statistical inference (SESI)** is a number $\alpha_{k,G} \in [0, 1]$ such that an $\alpha_{k,G}$ proportion of players take the action $A$ when each player draws $k$ independent observations from a Bernoulli distribution with probability of success $\alpha_{k,G}$, forms an estimate according to the inference procedure $G$, and best-responds to this estimate in choosing an action. We refer to $\alpha_{k,G}$ as the SESI proportion of degree $k$ with respect to the inference procedure $G$.

A SESI embodies two procedural constraints on players’ decision making. The first is informational: Players obtain data on the behavior of a small subset of players, e.g. because of time or other constraints. If players had been able to obtain data on the actions of all players, they would fully learn the equilibrium action profile, and the predictions of the model would coincide with those of Nash equilibrium, in which players are assumed to know this profile. The second constraint is cognitive: Players use only their data and statistical inference to estimate the equilibrium action profile. If players had thought strategically about how other players make inferences, the predictions of the model would again be identical to those of Nash equilibrium. The combination of the two constraints, however, implies different predictions than those of Nash equilibrium as we will see below.

Note that all players have the same sample size and use the same inference procedure in a SESI. We consider heterogeneity along these dimensions in Section 7.
2.4 Example: Consumption with negative externalities

Consider a situation in which the consumption of a good by one individual reduces the consumption utility of other individuals. For example, when attending an event such as a concert, the more people that attend the event, the more congested it is, and hence the less enjoyable the experience may be.

Let $\theta$ denote the idiosyncratic consumption benefit of the good and $\alpha^2$ the externality cost of consumption if a proportion $\alpha$ of the population consumes the good. An individual’s payoff is $\theta - \alpha^2$ when consuming the good, and 0 otherwise.

In order to decide whether to consume the good, the individual obtains a sample with $k$ observations and uses maximum likelihood to estimate $\alpha$. Thus, if the individual observes $j$ successes, his estimate is concentrated on $j/k$ and he consumes the good if his benefit $\theta$ exceeds the estimated externality cost $(j/k)^2$.

To solve for the SESI proportion $\alpha_{1,MLE}$, i.e., the SESI proportion for sample size 1 with respect to MLE, suppose that individuals sample from a Bernoulli distribution with success probability $\alpha$. Then:

- With probability $\alpha$, an individual draws a success. He estimates that the entire population consumes the good, and hence does not consume the good independently of $\theta$.

- With probability $1 - \alpha$, an individual draws a failure. He estimates that no one consumes the good, and hence he consumes the good independently of $\theta$.

Thus, the proportion $\alpha_{1,MLE}$ has to satisfy

$$\alpha = \alpha \cdot 0 + (1 - \alpha) \cdot 1.$$  

This equation has a unique solution at $\alpha_{1,MLE} = 1/2$.

For sample size 2, any SESI proportion $\alpha_{2,MLE}$ has to solve the equation

$$\alpha = \alpha^2 \cdot 0 + 2\alpha(1 - \alpha) \cdot (1 - 1/4) + (1 - \alpha)^2 \cdot 1.$$  

This is because:

- With probability $\alpha^2$, an individual draws two successes and does not consume the good.

- With probability $2\alpha(1 - \alpha)$, an individual obtains one success and one failure. He believes that half of the population consumes the good, and hence consumes the good if $\theta \geq (1/2)^2 = 1/4$. 
• With probability \((1 - \alpha)^2\), an individual obtains two failures and consumes the good independently of \(\theta\).

The equilibrium condition for sample size 2 is quadratic in \(\alpha\). It has a unique solution in \([0, 1]\) at \(\alpha_{2, MLE} = 1/2(\sqrt{17} - 3) \approx 0.56\). Note that \(\alpha_{1, MLE} < \alpha_{2, MLE}\).

Similarly, for \(k = 3\), any SESI has to solve the equation

\[
\alpha = \alpha^3 \cdot 0 + 3\alpha^2(1 - \alpha) \cdot (1 - 4/9) + 3\alpha(1 - \alpha)^2 \cdot (1 - 1/9) + (1 - \alpha)^3 \cdot 1.
\]

This cubic equation also has a unique solution in \([0, 1]\) at \(\alpha_{3, MLE} \approx 0.58\), which is larger than \(\alpha_{2, MLE}\).

If players use Beta estimation as in Example 2.3 instead of maximum likelihood, then the SESI predictions change for samples with two or more observations. For example, when the sample size is 2, a player who observes one success and one failure estimates that \(\alpha\) is distributed according to the Beta(1, 1) distribution, and hence that the expected externality is \(F_{2,1/2} = \int_0^1 z^2 dz = 1/3\).

The equation that characterizes the SESI proportion in this case is

\[
\alpha = \alpha^2 \cdot 0 + 2\alpha(1 - \alpha) \cdot (1 - 1/3) + (1 - \alpha)^2 \cdot 1.
\]

This equation also has a unique solution in \([0, 1]\) at \(\alpha_{2, Beta} \approx 0.54\). Note that \(\alpha_{2, Beta}\) is smaller than \(\alpha_{2, MLE}\).

### 3 Equilibrium characterization

This section establishes several properties of SESIs, which will serve as building blocks for solving applications in later sections. These properties include existence and uniqueness (Section 3.2), how SESIs relate to Nash equilibria (Section 3.3), and how they change as the sample size or the inference procedure changes (Sections 3.4 and 3.5 respectively). We prove these properties using the theory of Bernstein polynomials presented in Section 3.1.

#### 3.1 SESIs and Bernstein polynomials

The main technical tool that we use in equilibrium analysis is the theory of Bernstein polynomials.
**Definition** (Bernstein Polynomial). For a function \( v \) defined on the closed interval \([0, 1]\), the \( k \)-th order Bernstein polynomial of \( v \) is a function on \([0, 1]\) defined by

\[
\text{Bern}_k(x; v) \equiv \sum_{j=0}^{k} \binom{k}{j} x^j (1 - x)^{k-j} v(j/k).
\]

That is, \( \text{Bern}_k(x; v) \) is a polynomial of degree \( k \) in \( x \). It is the weighted average of the values of \( v \) at the \( k + 1 \) points \( \{0, \frac{1}{k}, \ldots, \frac{k-1}{k}, 1\} \), where the weight assigned to \( v(j/k) \) is the probability of obtaining \( j \) successes in \( k \) independent draws from a Bernoulli distribution with success probability \( x \). By definition \( \text{Bern}_k(0; v) = v(0) \) and \( \text{Bern}_k(1; v) = v(1) \).

**Lemma 1.** A number \( \alpha_{k,G} \) is a SESI proportion if and only if it solves the equation

\[
1 - \alpha = \text{Bern}_k(\alpha; F_k)
\]

where \( F_k \) is a function that assigns to every \( z \in [0,1] \) the value \( F_{k,z} \), i.e., \( F_k(z) = F_{k,z} \).

**Proof.** Suppose players obtain \( k \) independent draws from a Bernoulli distribution with success probability \( \alpha \) and use the inference procedure \( G \). The probability of observing \( j \) successes is

\[
\binom{k}{j} \alpha^j (1 - \alpha)^{k-j}.
\]

Conditional on observing \( j \) successes, all players with \( \theta \leq F_{k,j/k} \) take the action \( B \) where \( 0 \leq F_{k,j/k} \leq 1 \) because \( 0 \leq f(\alpha) \leq 1 \).

Thus, the fraction of players who observe \( j \) successes and take the action \( B \) is

\[
\binom{k}{j} \alpha^j (1 - \alpha)^{k-j} F_{k,j/k}.
\]

Summing over \( j \) yields the total measure of players taking the action \( B \), which is

\[
\sum_{j=0}^{k} \binom{k}{j} \alpha^j (1 - \alpha)^{k-j} F_{k,j/k} = \text{Bern}_k(\alpha; F_k).
\]

In equilibrium, this measure is equal to \( 1 - \alpha \).

As Lemma 1 indicates, analyzing properties of SESIs in our model is related to studying how Bernstein polynomials behave. The following properties of these polynomials, which they inherit from the function on which they operate, will be relevant in the analysis.

**Property 1.** If \( v \) is an increasing function in \( x \) then \( \text{Bern}_k(x; v) \) is an increasing function in \( x \) for any \( k \).
Property 2. If \(v\) is convex then \(\text{Bern}_k(x; v)\) is convex, and \(\text{Bern}_k(x; v) \geq \text{Bern}_{k+1}(x; v) \geq v(x)\) for any \(x\) with strict inequality for \(0 < x < 1\) if \(v\) is not linear.

Proofs of Properties 1 and 2 can be found in Phillips (2003). Another useful property that we prove directly is:

Property 3. Consider two inference procedures \(G\) and \(\tilde{G}\) such that \(G_{k,z}\) is a mean preserving spread of \(\tilde{G}_{k,z}\) for any sample \((k, z)\). Then \(\text{Bern}_k(\cdot; F_k) \geq \text{Bern}_k(\cdot; \tilde{F}_k)\).

Proof. By Jensen’s inequality and the convexity of \(f\), we obtain that \(F_{k,j/k} \geq \tilde{F}_{k,j/k}\) for any \(0 \leq j \leq k\). The property now follows directly from the definition of a Bernstein polynomial.

3.2 Equilibrium existence and uniqueness

Our first result is about existence and uniqueness of SESIs.

Theorem 1. There exists a unique SESI for any inference procedure and any sample size.

Proof of Theorem 1. Fix an inference procedure \(G\) and a sample size \(k\), and consider equation (1). The expression \(1 - \alpha\) on the left-hand side is strictly decreasing and continuous in \(\alpha\), it takes the value 1 at \(\alpha = 0\), and it takes the value 0 at \(\alpha = 1\).

The Bernstein polynomial \(\text{Bern}_k(\alpha; F_k)\) on the right-hand side has the following properties. First, it is a continuous function in \(\alpha\) on \([0, 1]\). Second, it increases in \(\alpha\). This is because the first-order stochastic dominance of \(G\) implies that \(F_k\) increases in \(z\), and the Bernstein polynomial of an increasing function is an increasing function (Property 1). Third, \(0 \leq \text{Bern}_k(0; F_k) < 1\) because \(0 \leq F_k(0) < F_k(1) \leq 1\) and \(F_k(0) = \text{Bern}_k(0; F_k)\). Similarly, \(0 < \text{Bern}_k(1; F_k) \leq 1\).

Thus, the function on the left-hand side and the Bernstein polynomial on the right-hand side cross exactly once on \([0, 1]\), which establishes existence and uniqueness.

There is also a unique Nash equilibrium (NE) proportion of players taking the action \(A\). To see this, consider a NE distribution of actions. Let \(\alpha_{NE}\) denote the proportion of players taking the action \(A\) in this NE, and let \(\theta_{NE}\) denote the type that satisfies \(\theta_{NE} = f(\alpha_{NE})\). All players with benefit above (below) \(\theta_{NE}\) strictly prefer to take the action \(A\) (\(B\)) and thus the proportion of players taking action \(A\) has to satisfy \(\alpha_{NE} = 1 - f(\alpha_{NE})\) or

\[1 - \alpha_{NE} = f(\alpha_{NE}).\] (2)

This equation has a unique solution on \([0, 1]\).
3.3 Relationship to Nash equilibrium

To compare the predictions of SESIs, which depend on the inference procedure and the sample size, to those of the unique NE, we restrict attention to unbiased inference procedures.

**Assumption 1** (Unbiased inference). An inference procedure \( G \) is **unbiased** if for any sample \((k, z)\) the expected value of the estimate \( G_{k,z} \) is equal to the sample mean, i.e.,

\[
\int_0^1 \alpha \, dG_{k,z}(\alpha) = z \quad \text{for every sample } (k, z).
\]

Assumption 1 is satisfied by the inference procedures in Examples 2.2, 2.3, and 2.4. However, the Bayesian Updating procedure in Example 2.1 does not satisfy this assumption. This is because whenever a player has a proper prior on \( \alpha \), and he updates using Bayes rule, the posterior mean of \( \alpha \) depends on both the prior mean and the sample mean. In particular, the posterior mean cannot coincide with the sample mean for all samples.

**Theorem 2.** The SESI proportion is strictly smaller than the NE proportion for any unbiased inference procedure \( G \) and any sample size \( k \) when the cost \( f \) is not linear.

**Proof of Theorem 2.** Fix a cost \( f \), an unbiased inference procedure \( G \), and a sample size \( k \). The unique NE proportion solves Equation (2) and the unique SESI proportion solves Equation (1). Because the left-hand side of both equations is identical, it suffices to prove — in order to establish the result — that the continuous and increasing functions on the right-hand side of both equations are ranked such that \( f(\alpha) \leq \text{Bern}_k(\alpha; F_k) \) with strict inequality for \( 0 < \alpha < 1 \) when \( f \) is not linear. This follows immediately from the equalities:

(i) \( \text{Bern}_k(\alpha; F_k) \geq \text{Bern}_k(\alpha; f) \) for every \( \alpha \in [0, 1] \)

(ii) \( \text{Bern}_k(\alpha; f) \geq f(\alpha) \) for every \( \alpha \in [0, 1] \) with strict inequality between 0 and 1 if \( f \) is not linear.

Inequality (ii) holds by Property 2. To verify inequality (i), recall that \( F_k = f \) for the maximum likelihood procedure, and that any unbiased inference procedure \( G \) is a mean preserving spread of the maximum likelihood procedure. Inequality (i) now follows from Property 3.

When the cost \( f \) is linear, the NE proportion and the SESI proportion coincide. This is because the linearity of \( f \) implies that \( F_k = f \) and so \( \text{Bern}_k(\alpha; F_k) = \text{Bern}_k(\alpha; f) \), and because the Bernstein polynomial of a linear function coincides with the function. This case is useful for
highlighting another important difference between the two solution concepts, which relates to
the selection of players who take each action.

In a NE, all players hold the same “correct” belief about the proportion of players taking
each action. Therefore, there is a positive sorting of players to actions in the sense that if some
player takes the action $A$, players with higher types also take the action $A$. In a SESI, players’
estimates differ from one another based on their sample. Players with a larger sample mean tend
to take the action $A$ less than players with a smaller sample mean leading to a weaker positive
sorting than in NE. It is even possible that there is no positive sorting of types to actions in a
SESI. This happens, for example, in the SESI with sample size 1 in Section 2.4, in which players’
actions depend only on their sample and not on their type.

How different are the predictions of the unique SESI and the unique NE? We return to the
example of Section 2.4 to demonstrate that the gap between the SESI and the NE proportions
can be large. Recall that $\alpha_{1, MLE}$ is 1/2 in this example. The unique NE proportion is the
solution to

$$1 - \alpha = \alpha^2,$$

which is about 0.62. Consider now the externality function $f(\alpha) = \alpha^n$. As $n$ increases, $\alpha_{1, MLE}$
does not change. However, $\alpha_{NE}$ increases and converges to 1 as $n \to \infty$.

### 3.4 Comparative statics: sample size

We proceed to evaluate how the SESI proportion changes as the sample size $k$ changes. To do
so, we make two additional assumptions on the inference procedure. The first is that players
become more confident in their estimate as the sample size increases.

**Assumption 2** (Noise Reduction). An inference procedure $G$ satisfies noise reduction if for
any two samples $(k, z)$ and $(\hat{k}, \hat{z})$ such that $\hat{k} > k$, the estimate $G_{k, z}$ is a mean preserving spread
of the estimate $G_{\hat{k}, \hat{z}}$.

Assumption 2 is satisfied by all the unbiased inference procedures in Section 2.2.

The second assumption is that fixing the sample size $k$, the inference procedure preserves the
shape of the function $f$ in the sense that the expected value of $f$ as a function of the proportion
of successes in the sample $z$ is convex whenever $f$ is convex.
Assumption 3 (Shape Preserving). An inference procedure $G$ is **shape preserving** if the expected cost function $F_k$ is convex whenever the cost function $f$ is convex.

The maximum likelihood procedure trivially satisfies Assumption 3 because $F_k = f$. A sufficient condition for Assumption 3 to hold is that, fixing $k$, the densities of the unbiased inference procedure $G_{k,z}(\alpha)$ are totally positive of degree 3 (TP3) in $(z, \alpha)$ (c.f. Jewitt (1988)).

The inference procedures in Examples 2.3 and 2.4 satisfy Assumption 3 because they belong to the exponential family, and exponential family densities are totally positive of every degree.

**Theorem 3.** If the inference procedure is unbiased, satisfies noise reduction, and preserves shape, then the SESI proportion of degree $k$ is strictly smaller than the SESI proportion of degree $k + 1$ when $f$ is not linear.

**Proof of Theorem 3.** Fix an inference procedure $G$. By Lemma 1, it suffices to prove that for all $\alpha \in [0, 1]$

$$Bern_k(\alpha; F_k) \geq Bern_{k+1}(\alpha; F_{k+1})$$

with strict inequality for $0 < \alpha < 1$. We establish this inequality in two steps.

**Step 1.** $Bern_k(\alpha; F_k) \geq Bern_k(\alpha; F_{k+1})$.

The convexity of $f$ together with noise reduction imply by Jensen’s inequality that $F_k(z) \geq F_{k+1}(z)$ for any $z \in [0, 1]$. The inequality follows.

**Step 2.** $Bern_k(\alpha; F_{k+1}) \geq Bern_{k+1}(\alpha; F_{k+1})$ with strict inequality for $0 < \alpha < 1$.

Because the inference procedure is shape preserving, $F_{k+1}$ is convex. The inequality now follows from Property 2.

By theorem 3, as $k$ grows the SESI proportion gets closer to the NE proportion. It converges to the NE proportion when the inference procedure is noiseless in the limit in the sense that for every $z$, $G_{k,z}$ converges, as $k$ tends to infinity, to the distribution that puts a unit mass on $z$. This is because in this case the function $F_k$ converges to the function $f$, and the Bernstein polynomial of any continuous function converges to the function as $k$ tends to infinity.

---

4A function $h(x, y)$ is totally positive of degree 3 if for any $x_1 < x_2 < x_3$ and $y_1 < y_2 < y_3$ the matrix $(h(x_i, y_j))$ has a non-negative determinant for each minor of size $\leq 3$. Total positivity of degree 2 (TP2) is the Monotone Likelihood Ratio Property, and is implied by TP3. TP3 ensures that the likelihood ratios increase sufficiently quickly to preserve convexity under integration.
Intuitively, as the sample size increases, each player is more likely to obtain a more accurate estimate of the proportion of players taking action $A$. As the sample size tends to infinity, this estimate converges to the exact proportion, and hence the SESI proportion converges to the NE proportion.\(^5\)

### 3.5 Comparative statics: inference procedure

The predictions of all unbiased inference procedures coincide when players draw a single observation. This is because unbiasedness implies that a player who draws a success must concentrate his estimate on $\alpha = 1$, and a player who draws a failure must concentrate his estimate on $\alpha = 0$. However, two unbiased inference procedures may differ in how they estimate $\alpha$ when players draw two or more observations.

**Definition** (Noisier Inference). The inference procedure $G$ is **noisier** than the inference procedure $\hat{G}$ if $G_{k,z}$ is a mean-preserving spread of $\hat{G}_{k,z}$ for any sample $(k, z)$.

Fix a sample size $k$. If $G$ is noisier than $\hat{G}$ then $\text{Bern}_k(\alpha; F_k) \geq \text{Bern}_k(\alpha; \hat{F}_k)$ by Property 3, and hence by Lemma 1 the SESI proportion with respect to $G$ is smaller than the SESI proportion with respect to $\hat{G}$. Thus, noisier inference implies a smaller proportion of players taking the action $A$.

The MLE procedure is the least noisy procedure among all unbiased procedures. Thus, for any given sample size, MLE provides the closest prediction to Nash equilibrium among all unbiased inference procedures. This is because by Theorem 2 the SESI proportions of all unbiased inference procedures are smaller than the NE proportions.

**Comment on concave costs.** The results of Sections 3.2–3.5 extend to the case of a concave cost function $f$ in a natural way. Theorem 1 continues to hold as it does not rely on the curvature of $f$. The comparison between the SESI proportion and the NE proportion in Theorem 2 is reversed, i.e. the SESI proportion is larger than the NE proportion, because the direction of the two relevant inequalities is reversed. As the sample size increases, the SESI

\(^5\)Of course, the SESI proportion does not converge to the NE proportion if the inference procedure $G$ has noise in the limit, i.e. there exists $\epsilon > 0$ such that the variance of $G_{k,z} \geq \epsilon$ for all $k$ and for all $z$ in an interval around the NE proportion. In this case, $F_k(z)$ is larger than $f(z) + \eta$ in this interval for some $\eta > 0$, and $\text{Bern}_k(\alpha; F_k)$ is even larger. Thus, the SESI proportion is bounded away from the NE proportion.
proportion decreases because the relevant inequilities in Theorem 3 are also reversed. Finally, the SESI proportion is larger for a noisier inference procedure.

4 Statistical inference in a competitive market

This section demonstrates how statistical inference can be incorporated to the theory of competitive markets.

A unit mass of producers face the inverse demand function $P(Q)$ for a good. Each producer has to decide whether to produce a unit of the good at an idiosyncratic cost $\theta$ to be sold at the market price. The market price depends the market supply $Q$, and so each producer needs to estimate $Q$ in order to decide whether to produce or not.

In a Nash equilibrium, producers best-respond to the same correct belief about the market supply $Q_{NE}$. Producers may arrive at this belief using the following thought process: The market supply has to be equal to the proportion of producers who decide to produce given this market supply, i.e., all producers with $\theta$ below the market price. Hence, $Q_{NE}$ has to solve

$$Q = P(Q).$$

On the other hand, when producers act as statisticians, each of them obtains a sample of the production decisions of other producers, uses statistical inference to estimate the market supply and the market price, and makes a decision based on the estimate. In a SESI, producers obtain their sample from a Bernoulli distribution with success probability governed by the market supply. Hence, the market supply $Q_{SESI}$ has to solve

$$Q = \sum_{j=0}^{k} \binom{k}{j} Q^j (1-Q)^{k-j} P_{k,j/k}$$

where $P_{k,j/k}$ is the estimated market price of producers who obtain the sample $(k,j/k)$ and use the estimate $G_{k,j/k}$ of the market supply $Q$.

To apply the results of Section 3 to solve for SESIs and compare them to NE in the current environment, we treat the action “don’t produce” as action $A$. Its benefit, which is the benefit of leisure, is $\theta$ and its cost is the foregone revenue from producing. Letting $N = 1 - Q$ denote

---

6The function $P(Q)$ summarizes a demand model with a unit mass of consumers whose valuations are distributed on the interval $[0,1]$ with a positive density function. Because consumers’ valuations are between 0 and 1, we can assume without loss of generality that $P(Q)$ is between 0 and 1.
the proportion of producers who do not produce, this cost is \( f(N) = P(1 - N) \). Thus, the utility of not producing becomes \( \theta - f(N) \) and the utility of producing becomes 0. Because \( P \) is decreasing in \( Q \), \( f \) is increasing in \( N \). And by definition, the functions \( f \) and \( P \) have the same convexity properties.

The following result is a corollary of Theorem 1.

**Corollary 1.** There exists a unique SESI market supply \( Q_{SESI} \) for every inference procedure and every sample size.

Because there is also a unique NE market supply, we can apply Theorem 2 and the market clearing condition to obtain the following.

**Corollary 2.** If the inverse demand function is convex in volume then there is over-supply of the good and a lower market price in the SESI than in the NE. On the other hand, if the inverse demand function is concave in volume then there is under-supply of the good and a higher market price in the SESI than in the NE.

## 5 Statistical inference and monopoly theory

This section demonstrates how statistical inference can be incorporated to monopoly theory, and discusses how the predictions of statistical inference differ from those of rational expectations in this setting.

Consumers sometimes have a preference for uniqueness when purchasing a product. This may be the case when the product conveys social status (e.g., jewelry) or when its consumption reflects freedom or independence (e.g., a designer clothing item). In such situations, an individual’s utility of consuming the good decreases the more individuals consume the good.

To model preference for uniqueness, we postulate that an individual has a consumption value \( \theta \) for the good, and a disutility of (-1) if the individual meets one or more other individuals who also have the good in \( t \) random encounters. Thus, if an \( \alpha \) proportion of the population purchases the good, the expected disutility of purchasing the good is \( 1 - (1 - \alpha)^t \), which we will approximate by \( 1 - e^{-\alpha t} \). The consumption utility of an individual is therefore \( u(\theta, \alpha) = \theta - (1 - e^{-\alpha t}) \).

A monopolist produces the good at zero marginal cost and sets a price \( p \) to maximize profit. Consumers observe the price and estimate the expected consumer demand for the good at this
price. They then decide whether to purchase the good. The monopolist knows consumers’ decision making process and takes this knowledge into account when pricing the good. We are interested in characterizing how the monopolist’s profit changes when consumers use statistical inference rather than rational expectations when estimating the demand for the product. Such a characterization may be informative about the monopolist’s incentives to release information about the expected demand for the product. To the extent that the monopolist chooses between releasing “accurate” information (in the sense that consumers’ best responses to this information will substantiate the information) or not releasing information and letting consumers form their own estimates, he would choose whether to release information or not based on the profit implications discussed below.

We first solve an example and then provide a general characterization.

**Example 5.1.** Suppose that \( t = 4 \) and that the monopolist sets a price \( p \) for the good. An individual who purchases the good obtains a utility of \( \theta - p - (1 - e^{-4\alpha}) \). Otherwise, his utility is 0. In order to decide whether to purchase the good, he has to estimate \( \alpha \).

Rational expectations require that \( \alpha \) is equal to the proportion of consumers with \( \theta \geq p + (1 - e^{-4\alpha}) \), i.e., \( \alpha \) has to satisfy \( \alpha = 1 - p - (1 - e^{-4\alpha}) \). The monopolist’s inverse demand is therefore \( p = e^{-4\alpha} - \alpha \). Solving the monopolist’s profit maximization problem, we obtain that the monopolist’s profit under rational expectations is 0.060 with an optimal quantity of 0.135 and an optimal price of 0.448.

Unbiased statistical inference from a single observation implies that when a consumer observes a success (i.e., someone who intends to purchase the good), he estimates \( \alpha \) to be 1 and hence his utility to be \( \theta - p - (1 - e^{-4}) \). For \( p \geq 0.02 \), this utility is negative for any \( \theta \leq 1 \), and thus no one will purchase the good in this case. When the consumer observes a failure, on the other hand, he estimates \( \alpha \) to be 0 and hence his utility to be \( \theta - p \). The proportion of individuals who purchase the good in this case is \( 1 - p \). Thus, at a price \( p \geq 0.02 \), the SESI demand has to satisfy \( \alpha = \alpha \cdot 0 + (1 - \alpha)(1 - p) \) and the inverse demand is \( p = 1 - \frac{\alpha}{1-\alpha} \). Solving the monopolist’s profit maximization problem, we obtain that the monopolist’s profit is 0.17 with an optimal quantity of 0.29 and an optimal price of 0.59. The monopolist’s profit is almost three-times larger than with rational expectations.

Unbiased statistical inference from two observations depends on how a consumer estimates demand when he observes one success and one failure. Suppose that he uses the Beta estimation
procedure and so has the estimate Beta(1, 1) following this sample. Then, when the consumer observes one success and one failure, he estimates the disutility from not being unique to be the expected value of $1 - e^{-4\alpha}$ with respect to Beta (1, 1), which is about 0.75. The proportion of individuals who purchase the product conditional on this sample is $\max\{0.25 - p, 0\}$. It is easy to verify that $p \leq 0.25$ is not optimal for the monopolist and thus the monopolist will cater only to those who observe two failures. The optimal profit would be smaller than with one observation at 0.14, the optimal quantity would also be smaller at 0.23 but the optimal price would be higher at 0.62.

More generally, we have the following comparative static result for unbiased inference procedures that satisfy noise reduction and preserve shape.

**Proposition 4.** *The monopolist’s profit when consumers use statistical inference is larger than when consumers have rational expectations. For a given inference procedure, the monopolist’s profit decreases in the sample size.*

**Proof.** Suppose the monopolist sets a price $p$ for the good. Fix an inference procedure $G$ and a sample size $k$. A consumer who observes $j$ successes purchases the good if his type $\theta$ is larger than $p + 1 - F_{k,j/k}$ where $F_{k,j/k}$ is the expected value of $e^{-\alpha t}$ according to $G_{k,j/k}$. Thus, conditional on observing $j$ successes, the proportion of consumers who purchase the good is $\max\{F_{k,j/k} - p, 0\}$. Demand at price $p$, $\alpha_{k,G}(p)$, is therefore determined by the solution to the equation

$$\alpha = \sum_{j=0}^{k} \binom{k}{j} \alpha^j (1 - \alpha)^{k-j} \max\{F_{k,j/k} - p, 0\}$$

which can be rewritten as

$$1 - \alpha = \sum_{j=0}^{k} \binom{k}{j} \alpha^j (1 - \alpha)^{k-j} \min\{1 + p - F_{k,j/k}, 1\}.$$ 

Because the function $e^{-\alpha t}$ is convex and decreasing, the function $F_k$ is convex and decreasing, and the function $1 + p - F_k$ is concave and increasing in $\alpha$. The function $\min\{1 + p - F_k, 1\}$ is therefore concave and increasing. We thus have an equation which is similar to equation 1 and the results of Section 3 continue to hold. That is, for every price $p$, the demand $\alpha_{k,G}(p)$ decreases in $k$ and is above the rational expectations demand, which is the $\alpha(p)$ that solves the equation $\alpha = 1 - e^{-\alpha t} - p$. We thus obtain a ranking of the demand functions the monopolist faces, and this ranking implies the ranking of the monopolist’s profit.

\[\square\]
A possible interpretation of Proposition 4 is that a monopolist in a market with preference for uniqueness has no incentive to inform consumers about his expectations regarding the demand for the good.

6 Statistical inference in two-sided markets

This section incorporates statistical inference to two-sided markets in which workers and firms need to decide whether to engage in costly search in order to create jobs.

There is a unit mass of workers, and each of them needs to decide whether to search for a job at an idiosyncratic search cost $\theta$. Workers’ costs are distributed uniformly on $[0, 1]$. There is also a unit mass of firms, and each of them needs to decide whether to post a vacancy at an idiosyncratic cost $\omega$. Firms’ costs are also distributed uniformly on $U[0, 1]$. If $\alpha$ workers search for jobs and $\beta$ firms search for workers, the number of jobs created is given by the matching function

$$m(\alpha, \beta) = \mu \alpha^x \beta^{1-x}$$

where $0 < \mu < 1$ is the matching friction that prevents full employment even if all workers and all firms participate in the market. As noted by Petrongolo and Pissarides (2001), “the matching function summarizes a trading technology between agents who place advertisements, read newspapers and magazines, go to employment agencies, and mobilize local networks that eventually bring them together into productive matches. The key idea is that this complicated exchange process is summarized by a well-behaved function that gives the number of jobs formed at any moment in time in terms of the number of workers looking for jobs, the number of firms looking for workers, and a small number of other variables. ... The stylized fact that emerges from the empirical literature is that there is a stable aggregate matching function of a few variables that satisfies the Cobb-Douglas restrictions with constant returns to scale in vacancies and unemployment.”

A worker who searches for a job finds one with probability $\frac{m(\alpha, \beta)}{\alpha}$ when $\alpha$ workers and $\beta$ firms participate in the market. Assuming that each match creates a surplus of 2 to be split equally between the worker and the firm, the expected utility of a worker who searches for a job is $\frac{m(\alpha, \beta)}{\alpha} - \theta$ where $\theta$ is the worker’s search cost. Similarly, the expected utility of a firm that posts a vacancy is $\frac{m(\alpha, \beta)}{\beta} - \omega$. Firms and workers who do not participate in the market obtain a
utility of 0. In order to decide whether to participate in the market, workers and firms estimate the market thickness $\alpha$ on the workers’ side and the market thickness $\beta$ on the firms’ side.

In a Nash Equilibrium, all workers and firms form the same belief about the market thickness $(\alpha_{NE}, \beta_{NE})$. This belief satisfies the rational expectations assumption whereby if workers and firms best-respond to the belief, actual market thickness is $(\alpha_{NE}, \beta_{NE})$. Thus, market thickness on the workers’ side $\alpha_{NE}$ has to satisfy that all workers with cost $\theta$ below the expected benefit of searching for a job $\frac{m(\alpha_{NE}, \beta_{NE})}{\alpha_{NE}} = \mu \alpha_{NE}^{x-1} \beta_{NE}^{1-x}$ decide to search for a job. Applying the same reasoning for firms, we obtain that the two equations that characterize the Nash Equilibrium participation are

$$\alpha = \mu \alpha^{x-1} \beta^{1-x}$$

$$\beta = \mu \alpha \beta^{-x}$$

We can rearrange the above equations as follows:

$$\alpha_{NE}(\beta) \equiv \left(\mu \beta^{1-x}\right)^{\frac{1}{1-x}}$$  \hspace{1cm} (3)

$$\beta_{NE}(\alpha) \equiv \left(\mu \alpha \right)^{\frac{1}{1+x}}$$  \hspace{1cm} (4)

We prove below that there is a unique Nash Equilibrium with positive participation.

We relax the rational expectations assumption as follows. We postulate that while each worker estimates market thickness on the workers’ side accurately, he uses statistical inference to estimate firms’ participation. This difference between workers’ reasoning ability regarding workers’ participation and firms’ participation aims to capture situations in which each worker has a good understanding of his side of the market, e.g. because he observes the decisions of many workers, but he struggles to understand firms’ decision making, e.g. because his access to information on firms’ decision making is limited. In order to get a better understanding of firms’ decision making, he collects data on the decisions of a few firms and uses statistical inference to estimate the market thickness on the firms’ side. We make an analogous assumption on firms’ reasoning abilities: each firm estimates the market thickness of the workers’ side accurately but samples a few workers in order to learn about workers’ participation. Formally,

**Definition** (SESI in a two-sided market). A **SESI in a two-sided markets** is a pair of numbers $(\alpha_{k,Gw}, \beta_{m,Gf}) \in [0, 1] \times [0, 1]$ such that
\[ \alpha_{k,G^w} \] workers search for a job when each worker (1) obtains \( k \) random draws about firms’ behavior from a Bernoulli distribution with probability of success \( \beta_{m,G^f} \), and (2) best-responds to the estimate he forms according to the inference procedure \( G^w \) and \( \alpha_{k,G^w} \); and

\[ \beta_{m,G^f} \] firms search for workers when each firm (1) obtains \( m \) random draws about workers’ behavior from a Bernoulli distribution with probability of success \( \alpha_{k,G^w} \), and (2) best-responds to the estimate it forms according to the inference procedure \( G^f \) and \( \beta_{m,G^f} \).

To develop the equations that characterize SESIs in this setting, fix a pair of unbiased inference procedure \((G^w, G^f)\) that preserve shape and satisfy noise reduction. Let \( M(\alpha, (k, z)) \) denote the expected matching probability of a worker who estimates that \( \alpha \) workers are in the market, obtains the sample \((k, z)\) regarding firms’ behavior, and uses the inference procedure \( G^w \). Let \( M((m, z'), \beta) \) denote the analogous expected matching probability of a firm that observes the sample \((m, z')\) of workers’ behavior. Then, the proportion of workers who participate in the market conditional on observing the sample \((k, z)\) is equal to the expected matching probability \( M(\alpha, (k, z)) \) and similarly the proportion of firms is \( M((m, z'), \beta) \). The probability that a worker observes the sample \((k, z)\) with \( z = j/k \) is \( \binom{k}{j} \beta^j (1 - \beta)^{k-j} \) and the probability that a firm observes the sample \((m, z')\) with \( z' = i/m \) is \( \binom{m}{i} \alpha^i (1 - \alpha)^{m-i} \). The equations that characterize any SESI are therefore

\[
\alpha = \sum_{j=0}^{k} \binom{k}{j} \beta^j (1 - \beta)^{k-j} M(\alpha, (k, j/k))
\]

\[
\beta = \sum_{i=0}^{m} \binom{m}{i} \alpha^i (1 - \alpha)^{m-i} M((m, i/m), \beta).
\]

The expected matching probability of workers \( M(\alpha, (k, z)) \) satisfies \( M(\alpha, (k, z)) = \mu \alpha^{x-1} M^w(k, z) \) where \( M^w(k, z) \) is the expected value of \( \beta^{1-x} \) with respect to the estimate \( G^w_{k, z} \). Rewriting the expected matching probability of firms similarly, we can rearrange the above equations into

\[
\alpha_k(\beta) \equiv \left( \mu \sum_{j=0}^{k} \binom{k}{j} \beta^j (1 - \beta)^{k-j} M^w(k, j/k) \right)^{1/(1-x)}
\] \hspace{1cm} (5)

\[
\beta_m(\alpha) \equiv \left( \mu \sum_{i=0}^{m} \binom{m}{i} \alpha^i (1 - \alpha)^{m-i} M^f(m, i/m) \right)^{1/(1-x)}.
\] \hspace{1cm} (6)
While the environment and the equations characterizing SESIs in two-sided markets are different from those of Sections 2 and 3, we can still use the theory of Bernstein polynomials to characterize SESIs in two-sided markets.

**Theorem 5.** Fix a pair of unbiased inference procedures $G^w$ and $G^f$ that preserve shape and satisfy noise reduction, a sample size $k$ for workers, and a sample size $m$ for firms. Then, there exists a unique SESI with positive employment. In this positive SESI, market thickness on the workers’ and the firms’ sides, and hence employment, are smaller than in the unique Nash equilibrium with positive employment. Moreover, as $k$ or $m$ increase, market thickness on both sides of the market as well as employment in the positive SESI increase.

**Proof.** The function $\alpha_k(\beta)$ is strictly concave and strictly increasing in $\beta$. This is because (1) $\beta^{1-x}$ is concave and strictly increasing in $\beta$, hence (2) $M^w(k, z)$ is concave and strictly increasing in $z$ because $G^w$ is an inference procedure that preserves shape, hence (3) the Bernstein polynomial of $M^w(k, z)$, which is the expression in parenthesis in equation 5, is concave and strictly increasing, hence this expression raised to the power of $\frac{1}{2-x} < 1$ is strictly concave and strictly increasing in $\beta$. A similar argument establishes that $\beta_m(\alpha)$ is strictly concave and strictly increasing in $\alpha$, which implies that its inverse function $\hat{\alpha}_m(\beta)$ is strictly convex and strictly increasing in $\beta$. Now, $\beta$ is part of a SESI if and only if it is a point in which the functions $\alpha_k(\beta)$ and $\hat{\alpha}_m(\beta)$ intersect. Because $\alpha_k(\beta)$ is strictly concave and $\hat{\alpha}_m(\beta)$ is strictly convex and both are increasing, they intersect in at most one positive point. Since the values of both functions at $\beta = 0$ is zero, and at $\beta = 1$, $\alpha_k(\beta) < \hat{\alpha}_m(\beta)$, we obtain that such a point exists. (A similar argument can be used to establish that there is a unique Nash equilibrium with positive employment.)

Arguments similar to the ones developed in Theorems 2 and 3 can be used to show that the function $\alpha_k(\beta)$ lies below the function $\alpha_{NE}(\beta)$ and that the function $\alpha_{k+1}(\beta)$ lies between these two functions for any $0 < \beta < 1$. Similarly, the function $\hat{\alpha}_m(\beta)$ lies above the function $\hat{\alpha}_{NE}(\beta)$ and the function $\hat{\alpha}_{m+1}(\beta)$ lies between these two functions. This yields the ranking of the equilibria.

The magnitude of under-employment in a SESI can be quite substantial. The following example illustrates.

**Example 6.1.** Suppose that $x = 1/2$. Then a Nash equilibrium is characterized by the equations $\alpha = \mu(\beta/\alpha)^{1/2}$ and $\beta = \mu(\alpha/\beta)^{1/2}$. In the unique Nash equilibrium with positive employment,
\( \alpha = \beta = \mu \) and employment is \( \mu^2 \). On the other hand, the unique positive SESI with unbiased inference procedures and sample sizes 1 is characterized by the equations \( \alpha = \beta \mu (1/\alpha)^{1/2} \) and \( \beta = \alpha \mu (1/\beta)^{1/2} \). In this SESI, participation is smaller by a factor of \( \mu \) at \( \alpha = \beta = \mu^2 \) and employment is smaller by the same factor, and is \( \mu^3 \).

7 Conclusion

This paper made modest progress in incorporating statistical inference into games. We postulated that players act as statisticians. They collect information about the actions of a few other players and use statistical inference procedures such as Maximum Likelihood or Beta estimation to estimate other players’ actions. In particular, they do not “think strategically” about how other players make inferences or arrive at their own actions. In order to “close” the model, we postulated that the data generating process from which players obtain information is representative of the distribution of players’ actions. A solution to this model, which depends on players’ sample size and inference procedure, is called a SESI. We developed tools for the analysis of SESIs, and used these tools to study the predictions of SESIs in several settings including competitive markets, monopoly theory, and two-sided markets.

The analysis so far assumed that all players obtain the same number of observations and use the same inference procedure. We conclude by showing that the analysis and the results are robust to heterogeneity along these dimensions. For simplicity, we focus on the setting of Sections 2 and 3.

Heterogeneous sample size. Heterogeneity in sample size may arise due to differences in individuals’ access to information or due to differences in the ability to recall or process information. To model such heterogeneity, let \( \gamma_i \) denote the proportion of individuals who obtain a sample of size \( i \) with \( 1 \leq i \leq K \) and \( \sum_{i=1}^{K} \gamma_i = 1 \) and let \( \gamma = (\gamma_1, \ldots, \gamma_K) \). Fix a convex cost function \( f \) and an unbiased inference procedure \( G \) that preserves shape and satisfies noise reduction.

A SESI of degree \( \gamma = (\gamma_1, \ldots, \gamma_K) \) with respect to the inference procedure \( G \) is a number \( \alpha_{\gamma,G} \in [0,1] \) such that an \( \alpha_{\gamma,G} \) proportion of players take the action \( A \) when, for every \( 1 \leq i \leq K \) a proportion \( \gamma_i \) of players draws \( i \) independent observations from a Bernoulli distribution with probability of success \( \alpha_{\gamma,G} \), and each player forms an estimate using the inference procedure \( G \) and best-responds to this estimate in choosing an action.
Similarly to Lemma 1, any SESI proportion $\alpha_{\gamma,G}$ is a solution to

$$1 - \alpha = \sum_{i=1}^{K} \gamma_i \text{Bern}_i(\alpha; F_i)$$

where $F_i$ is a function that assigns the value $F_{i,z}$ to every $z \in [0, 1]$.

The left-hand side of Equation 7 is identical to the left-hand side of Equation 1. The function on the right-hand side of Equation 7 is a finite convex combination of the functions on the right-hand side of Equation 1. It is easy to verify that this convex combination inherits the relevant properties of the functions on which it operates that we used in the proofs of Theorems 1 and 2. It therefore follows that (i) there exists a unique SESI proportion for any $\gamma$ and any $G$, and that (ii) this SESI proportion is smaller than the NE proportion.

In order to rank two SESI proportions, $\alpha_{\gamma,G}$ and $\alpha_{\hat{\gamma},G}$, we treat $\gamma$ and $\hat{\gamma}$ as discrete distributions over the values $1, \ldots, K$. Then, $\alpha_{\gamma,G}$ is strictly smaller than $\alpha_{\hat{\gamma},G}$ if $\gamma$ is first-order stochastically dominated by $\hat{\gamma}$.

To see why, recall that we established in Theorem 3 that $\text{Bern}_i(\alpha; F_i) \geq \text{Bern}_{i+1}(\alpha; F_{i+1})$ with strict inequality for $0 < \alpha < 1$. This implies that for every $0 < \alpha < 1$ the right-hand side of Equation 7 with respect to $\gamma$ is larger than with respect to $\hat{\gamma}$ by the first-order stochastic dominance, and the two expressions are equal in the end points. Therefore, the SESI proportion $\alpha_{\gamma,G}$ is smaller.

**Heterogeneous inference procedures.** Heterogeneity in inference procedures may arise for various reasons such as a preference for a particular estimation procedure or different levels of confidence in the estimation. To model such heterogeneity, let $G^1, \ldots, G^M$ denote the $M$ inference procedures used by players and let $\gamma_i$ denote the proportion of individuals who use the inference procedure $G^i$.

A SESI of degree $k$ with respect to the distribution $\gamma = (\gamma_1, \ldots, \gamma_M)$ of inference procedures $G = (G^1, \ldots, G^M)$ is a number $\alpha_{k,\gamma} \in [0, 1]$ such that an $\alpha_{k,\gamma}$ proportion of players take the action $A$ when each player draws $k$ independent observations from a Bernoulli distribution with probability of success $\alpha_{k,\gamma}$, for every $1 \leq i \leq M$ a proportion $\gamma_i$ of players form an estimate according to the inference procedure $G^i$, and each player best-responds to his estimate in choosing an action.

Assume that every $G^i$ is unbiased, preserves shape, and satisfies noise reduction. Fix a convex cost function $f$ and a sample size $k$, and let $F^i_k$ denote the function that assigns to any $z \in [0, 1]$ the expected value of $f$ with respect to $G^i$ conditional on the sample $(k, z)$.  

27
Similarly to Lemma 1, any SESI proportion $\alpha_{k,\gamma}$ has to solve

$$1 - \alpha = \sum_{i=1}^{M} \gamma_i \text{Bern}_k(\alpha; F_i^k).$$

(8)

Arguments similar to the ones used above and in Section 3 can now be used to establish that (i) there exists a unique SESI proportion for every sample size and every collection of inference procedures, (ii) this SESI proportion is smaller than the NE proportion, and (iii) the SESI proportion $\alpha_{k,\gamma}$ increases in $k$.

Thus, the results of Section 3 extend to heterogeneous samples and heterogeneous inference procedures. A challenge for future research is to extend the analysis and results to settings with dynamic or correlated sampling and to games with more than two actions.

References


