HOTELLING COMPETITION AND THE GAMMA DISTRIBUTION

BASSEL TARBUSH

ABSTRACT. I show that in any symmetric Nash equilibrium of Hotelling's pure location game, as the number of firms becomes large, the limiting distribution of market shares converges to a Gamma$(2, 2)$ distribution. Remarkably, this is true regardless of the distribution of consumers. The proof for the invariance of the limiting market share distribution builds on results from the theory of statistical spacings.

1. INTRODUCTION

Consumers, who buy only from the nearest firm, are distributed on the unit interval and firms independently and simultaneously choose their location on the interval to maximise their market share. This pure location game is a simplified version of Hotelling's (1929) seminal location model that ignores the firms' pricing decision. The game has been hugely influential in the social sciences and it has been applied in a number of different contexts including general spatial competition (Boppana et al., 2016; Núñez and Scarsini, 2017), competition between firms through product differentiation (Gabszewicz and Thisse, 1992), political party competition (Xefteris, 2016; Ronayne, 2016), and contests between professional forecasters (Laster et al., 1999; Ottaviani and Sørensen, 2006; Ewerhart, 2015).

In contrast to the existing literature, which has focused primarily on the equilibrium location choices of the firms, I study the resulting market structure (i.e. the distribution of market shares). In any symmetric Nash equilibrium of Hotelling's pure location game (in which all firms employ the same location strategy), as the number of firms becomes

large, each firm chooses its location at random according to a distribution that is equal to the distribution of consumers (Osborne and Pitchik, 1986). The game therefore features a stochastic process (via random location choices) that induces a distribution of market shares. I show that in any symmetric equilibrium of the game, as the number of firms becomes large, the distribution of market shares tends to a Gamma$(2, 2)$ distribution *regardless of the distribution of consumers.*

The main result of this paper therefore pins down the limiting market structure of the pure location game, which turns out to be invariant to the distribution of consumers.[1] Methodologically, I derive the main result by building on the theory of statistical "spacings", which are differences between order statistics.[2] Weiss (1955) and Pyke (1965) characterise the limiting distribution of spacings of order one (first-order differences between order statistics with an arbitrary parent distribution). I extend their result to spacings of order $m \geq 1$. The invariance of the limiting market share distribution comes from the fact that, when the number of firms is large, the distribution from which the location choices of the firms are drawn converges to the distribution of consumers, and this implies that the market share of any firm becomes proportional to the second-order difference between uniform order statistics regardless of the distribution of consumers. I then apply the general result on spacings of order m to derive the limiting distribution of these second-order differences.[3] Of course, the main result applies not only in the context of competition among firms, but also in the context of political party competition and of professional forecasting, and I discuss these cases in the paper.

The results of this paper can be applied to examine the limiting market structure in non-trivial extensions of the pure location game. In the classic formulation of the game, each firm i sells only to the consumers closest to i who are located between it and its nearest neighbours to the left and to the right, and it is under this formulation that the limiting distribution is Gamma$(2, 2)$. As an example, I consider a variant of the game in which each firm i sells to all consumers closest to i who are located between it and its *second* nearest neighbours to the left and to the right. The markets of the firms overlap and competitive pressure is reduced in this case. When the distribution of consumers is uniform, a straightforward application of the results of this paper shows that the limiting distribution of market shares is Gamma$(4, 2)$.

---

[1] Neven (1986) and Ansari et al. (1994) study the equilibria of Hotelling's game when a pricing decision is included for the case of a non-uniform distribution of consumers. I allow for any distribution of consumers in this paper. See Fournier (2016) and Núñez and Scarsini (2017) and references therein for papers that have considered arbitrary distributions of consumers in games of pure location.

[2] In a completely different application, Gabaix and Landier (2008) also appeal to results on spacings for the analysis of their economic model.

[3] I employ the result on spacings of order m in a heuristic proof of the main result. I provide a distinct formal proof for the main result in the appendix.

Such a distribution exhibits less inequality than a Gamma$(2, 2)$ distribution according to the Lorenz order (Kleiber and Kotz, 2003).

2. The model

I present the standard pure location model here and state a useful result regarding symmetric equilibria in the limit as the number of firms becomes large that is due to Osborne and Pitchik (1986). The exposition here closely follows theirs.

Consumers are distributed on the unit interval according to a cumulative distribution function $C \in \mathcal{C}^2$ (i.e. C is twice continuously differentiable) with full support. The location of each firm $i \in \{1, ..., n\}$ on the unit interval is determined by a random variable X_i that is drawn according to a distribution function $L_{i,n}$. Each firm i chooses its location strategy $L_{i,n}$ to maximise its expected market share. The strategies are chosen simultaneously and the random variables X_i are independent across i. For any realisation of locations $\mathbf{x} := (x_1, ..., x_n)$ resulting from the strategy profile $(L_{1,n}, ..., L_{n,n})$ the market share of firm i is given by

$$
s_i(\mathbf{x}) := \begin{cases} [C(\rho_i(\mathbf{x})) - C(\lambda_i(\mathbf{x}))]/\mu_i(\mathbf{x}) & \text{if } \mathtt{left}_i(\mathbf{x}) \neq \emptyset, \mathtt{right}_i(\mathbf{x}) \neq \emptyset \\ C(\rho_i(\mathbf{x}))/\mu_i(\mathbf{x}) & \text{if } \mathtt{left}_i(\mathbf{x}) = \emptyset, \mathtt{right}_i(\mathbf{x}) \neq \emptyset \\ [1 - C(\lambda_i(\mathbf{x}))]/\mu_i(\mathbf{x}) & \text{if } \mathtt{left}_i(\mathbf{x}) \neq \emptyset, \mathtt{right}_i(\mathbf{x}) = \emptyset \\ 1/\mu_i(\mathbf{x}) & \text{if } \mathtt{left}_i(\mathbf{x}) = \emptyset, \mathtt{right}_i(\mathbf{x}) = \emptyset \end{cases},
$$

where $\mathtt{left}_i(\mathbf{x})$ and $\mathtt{right}_i(\mathbf{x})$ are the locations of the firms (not including i) that are to the left and right of i's location respectively, $l_i(\mathbf{x})$ is the location of the firm closest to i in $\mathtt{left}_i(\mathbf{x})$, $r_i(\mathbf{x})$ is the location of the firm closest to i in $\mathtt{right}_i(\mathbf{x})$, and $\lambda_i(\mathbf{x}) := (x_i + l_i(\mathbf{x}))/2$ and $\rho_i(\mathbf{x}) := (x_i + r_i(\mathbf{x}))/2$. The number $\mu_i(\mathbf{x})$ is the number of firms (including i) located at x_i.

Characterising the full set of equilibria of this game for a finite number of firms is notoriously difficult (e.g. see Ewerhart, 2015). Eaton and Lipsey (1975) show that when the distribution of consumers is differentiable and C' has finitely many maxima, the game possesses no pure strategy equilibria when the number of firms becomes large.[4] On the other hand, the results of Dasgupta and Maskin (1986a,b) guarantee the existence of symmetric mixed strategy equilibria, i.e. $L_{i,n} = L_n$ for all i, for any n. Following many other papers in the literature, I therefore focus only on profiles of the form $(L_n, ..., L_n)$ where L_n is some non-degenerate distribution.

---

[4]Firm i's location strategy is said to be pure if $L_{i,n}$ is degenerate, and it is said to be mixed otherwise.

The expected market share of a firm i that chooses to locate at $x \in [0, 1]$ when every other firm chooses its location randomly according to the distribution L_n is given by

$$(1) \quad \pi_i(x) := (n-1) \int_x^1 (1 - L_n(y))^{n-2} \left[C\left(\frac{x+y}{2}\right) - 0 \right] dL_n(y)$$

$$+ \sum_{k=1}^{n-2} \binom{n-1}{k} k(n-k-1)$$

$$\int_0^x \int_x^1 L_n(z)^{k-1} (1 - L_n(y))^{n-k-2} \left[C\left(\frac{x+y}{2}\right) - C\left(\frac{x+z}{2}\right) \right] dL_n(z) dL_n(y)$$

$$+ (n-1) \int_0^x L_n(y)^{n-2} \left[1 - C\left(\frac{x+y}{2}\right) \right] dL_n(y).$$

The expression above accounts for all the possible events for firm i when it locates at x. Namely, all other firms locate to the right of x; $k = 1, ..., n-2$ locate to the left of x and $n - k - 1$ locate to the right of x; and all other firms locate to the left of x.

In a symmetric mixed strategy Nash equilibrium of the game, firm i must be indifferent over all locations x in the support of L_n. This therefore defines a functional equation for which the solution is a distribution L_n. Solving this functional equation for finite n and an arbitrary distribution of consumers C is analytically intractable (Osborne and Pitchik, 1986; Ewerhart, 2015). Yet Osborne and Pitchik (1986) proved the following limiting result.

Proposition 1 (Osborne and Pitchik, 1986). *Suppose that for each n the profile $(L_n, ..., L_n)$ is a symmetric mixed strategy Nash equilibrium and $L_n \in \mathcal{C}^2$. If $L_n \to L$, $L_n' \to L'$, $L_n'' \to L''$ uniformly for some $L \in \mathcal{C}^2$, then $L = C$.*

This result essentially states that in any symmetric equilibrium of the pure location game each firm plays a mixed strategy that is equal to the distribution of consumers when the number of firms is large. To be clear, this result holds *in the limit*: a profile $(L, ..., L)$ in which $L = C$ may not be an equilibrium for a finite number of firms (e.g. see Shaked, 1982).

The fact that the location choice of firms mimics the distribution of consumers in the limit is a result that is robust across different location models (Laster et al., 1999; Ottaviani and Sørensen, 2006; Núñez and Scarsini, 2017). Osborne and Pitchik (1986) explain the intuition for the limit result as follows: suppose for sake of argument that consumers are distributed uniformly but that the limiting distribution of firms is not uniform. Then any location y at which the density of consumers exceeds the density of firms will yield a larger market share than a location z at which the density of firms exceeds that of consumers. It follows that a firm would prefer to locate at y than at z, which violates the equilibrium condition that the firm be indifferent across all locations.
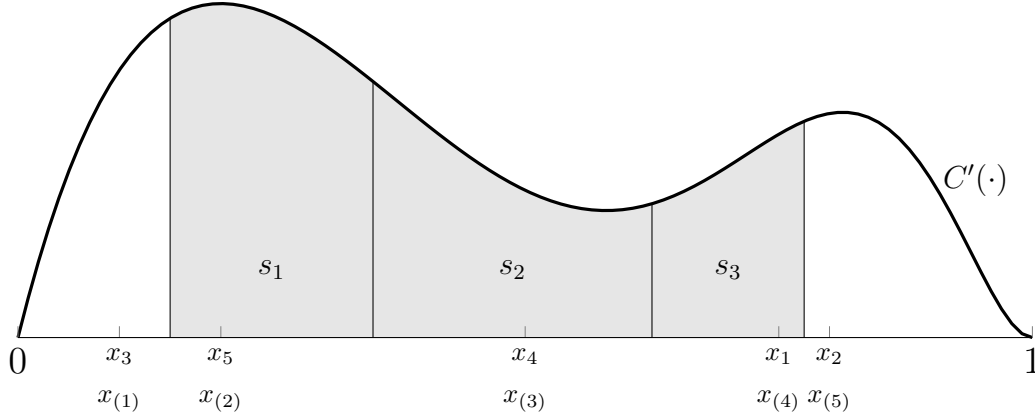
FIGURE 1. The realisations of X_i, $X_{(i)}$, and S_i are denoted by x_i, $x_{(i)}$, and s_i respectively. Here, we see a possible realisation of $n = 5$ locations. The share s_2, for example, is the area under the density of consumers $C'(\cdot)$ between $(x_{(3)} + x_{(2)})/2$ and $(x_{(4)} + x_{(3)})/2$.

The limiting result of Osborne and Pitchik (1986) is extremely useful because it tells us how the locations are chosen along the unit interval, but it does not tell us what the resulting limiting distribution of market shares will be.

3. THE MAIN RESULT

For the class of equilibria studied in Osborne and Pitchik (1986), the firms' location choices generate a sample $\mathcal{X} := \{X_1, ..., X_n\}$ of random variables in which each X_i is drawn independently from a distribution $L_n \in \mathcal{C}^2$. Denote the ith order statistic of the sample by $X_{(i)}$.[5] Without any loss of generality relabel the firms such that the market share of firm $i + 1$ is given by

$$(2) \qquad S_i := C\left(\frac{X_{(i+2)} + X_{(i+1)}}{2}\right) - C\left(\frac{X_{(i+1)} + X_{(i)}}{2}\right),$$

where $i \in \{1, ..., n - 2\}$.[6] Indeed, if ties are ignored (which have measure zero here) the market share of firm i is simply the integral of the density of consumers evaluated between the mid-points between firm i's location and the locations of its nearest neighbours.

An illustration of the firms' market shares is given in Figure 1 for a particular realisation of firm locations. Notice that there are $n - 2$ rather than n market shares: the market shares of the firms at each end point of the unit interval are ignored, but ignoring these two data points becomes irrelevant when n is large.

---

[5]That is, $X_{(i)}$ is the ith smallest value in $\mathcal{X}$. So, for example, $X_{(1)} = \min \mathcal{X}$ and $X_{(n)} = \max \mathcal{X}$.
[6]The firms were labelled by the unordered variables $X_1, ..., X_n$ but are now identified by the ordered variables $X_{(1)}, ..., X_{(n)}$.

We will be concerned with the empirical distribution of the variables S_i. Specifically, let us consider the variable $G_n(x) := \frac{1}{n-2} \sum_{i=1}^{n-2} \mathbf{1}(nS_i \leq x)$. The summation is the number of market shares (scaled by n) that do not exceed some value $x \geq 0$.[7] Since this value is divided by $n-2$, the variable $G_n(x)$ is the empirical fraction of market shares (scaled by n) that do not exceed x. The main result of the paper is stated below.

Theorem 1. *Suppose that for each n the profile $(L_n, ..., L_n)$ is a symmetric mixed strategy Nash equilibrium and $L_n \in \mathcal{C}^2$. If $L_n \to L$, $L'_n \to L'$, $L''_n \to L''$ uniformly for some $L \in \mathcal{C}^2$, then as $n \to \infty$,*

$$\sup_{x \geq 0} \left| G_n(x) - G(x) \right| \xrightarrow{p} 0,$$

where

$$G(x) := 1 - (2x + 1) \exp\{-2x\}.$$

In words, the result essentially states that in any symmetric equilibrium (of the broad class covered in Osborne and Pitchik, 1986) the distribution of market shares approaches a Gamma$(2, 2)$ distribution. Indeed, $G_n(x)$, the fraction of market shares (scaled by n) not exceeding x, approaches the cumulative distribution function $G(x)$ when n is taken to be arbitrarily large and the density $G'(x) = 4x \exp\{-2x\}$ corresponds to a Gamma$(2, 2)$ distribution. With some hand-waving one could also read the result as stating that the distribution of market shares (*not* scaled by n) is well-approximated by a Gamma$(2n, 2n)$ distribution when n is large.

Notice that the distribution of consumers C does not appear in the statement of Theorem 1. The limit result of Osborne and Pitchik (1986) tells us that the firms' location choices will tend to the distribution of consumers, but the striking feature of Theorem 1 is that, when this happens, the distribution of market shares remains Gamma$(2, 2)$ *regardless* of the distribution of consumers. In other words, the limiting distribution of market shares is invariant to the distribution of consumers.

3.1. **Simulations.** To illustrate Theorem 1 numerically one may want to computationally generate the equilibrium strategy L_n for each n and let n become arbitrarily large. Doing this would allow us to track the evolution of both the equilibrium strategy as well as the distribution of market shares as n increases. However, generating the equilibrium strategy L_n for each n is complex: as explained in section 2, this would require solving a functional equation involving (1). For the specific case in which the distribution of consumers is uniform, Ewerhart (2015) derived a system of differential equations that characterises the "doubly" symmetric equilibrium strategy profile for finite n and used the system to numerically compute

---

[7]The market shares obviously go to zero as $n \to \infty$ so the scaling is required to prevent the distribution from becoming degenerate when the limit is taken.

the equilibrium profile for small values of n.[8] Fully characterising symmetric equilibrium profiles $(L_n, ..., L_n)$ for finite n and a general distribution of consumers remains an open question.

For the purpose of illustrating Theorem 1, *which is about the limit*, I do not attempt explicit computation of L_n for each n. Rather, I rely on the result of Osborne and Pitchik (1986): since L_n becomes C in the limit, the following is a straightforward corollary of Theorem 1.

Corollary 1. *Suppose that for each n the profile $(L_n, ..., L_n)$ is equal to $(C, ..., C)$, then as $n \to \infty$, $\sup_{x \geq 0} |G_n(x) - G(x)| \xrightarrow{p} 0$.*

That is, when all firms play the *limiting* equilibrium strategy C, the distribution of market shares becomes $\text{Gamma}(2, 2)$ as n becomes large. To illustrate the content of Theorem 1 numerically, I therefore let all firms choose their location directly according to the distribution of consumers (i.e. for each n, L_n is set equal to the *limiting* symmetric equilibrium strategy C) and I then choose n to be a large number.

The results of the numerical illustration are shown in Figure 2. I set the distribution of consumers to be uniform on $[0, 1]$ (panel (a_1)) and let $n = 1000$ firms each draw their location independently from this distribution. The histogram of the resulting market shares (scaled by n) is shown in panel (a_2) in grey.[9] The density of a $\text{Gamma}(2, 2)$ distribution is shown in black and the fit is good. To illustrate the invariance of the limiting market share distribution, in a separate experiment I set the distribution of consumers to be a $\text{Beta}(7, 5)$ (panel (b_1)) and let the firms draw their locations from this distribution. The resulting market share distribution is shown in panel (b_2). Once again, a $\text{Gamma}(2, 2)$ distribution offers a good fit.

Further simulations relating to this section and to other results in the paper are presented in appendix B.

3.2. **Preliminary discussion.** Theorem 1 is relevant in the various different contexts in which Hotelling's pure location game has been applied. These include competition across firms, political party competition, and contests between professional forecasters.

The model presented in section 2 describes competition among firms. The unit interval can, for example, be interpreted either as physical space such as a street or as the product space (Gabszewicz and Thisse, 1992). Firms choose their location (on the street or in the product space) to maximise their market share. If the firms play the same strategy, then Theorem 1 tells us that the distribution of scaled market shares will be $\text{Gamma}(2, 2)$ when the number

---

[8]The profile is doubly symmetric if all firms use the same location strategy L_n and L_n is reflected at the mid-point of the unit interval, i.e. $L_n(1 - x) = 1 - L_n(x)$ for all $x \in [0, 1]$.

[9]The distribution has a long tail so the figure is zoomed in on the distribution for scaled shares between 0 and 10. The code for the simulations is available at http://users.ox.ac.uk/~scat3580/Hotelling.zip.

(a_1) U$(0,1)$ distribution of consumers (a_2) Resulting market share distribution

(b_1) Beta$(7,5)$ distribution of consumers (b_2) Resulting market share distribution
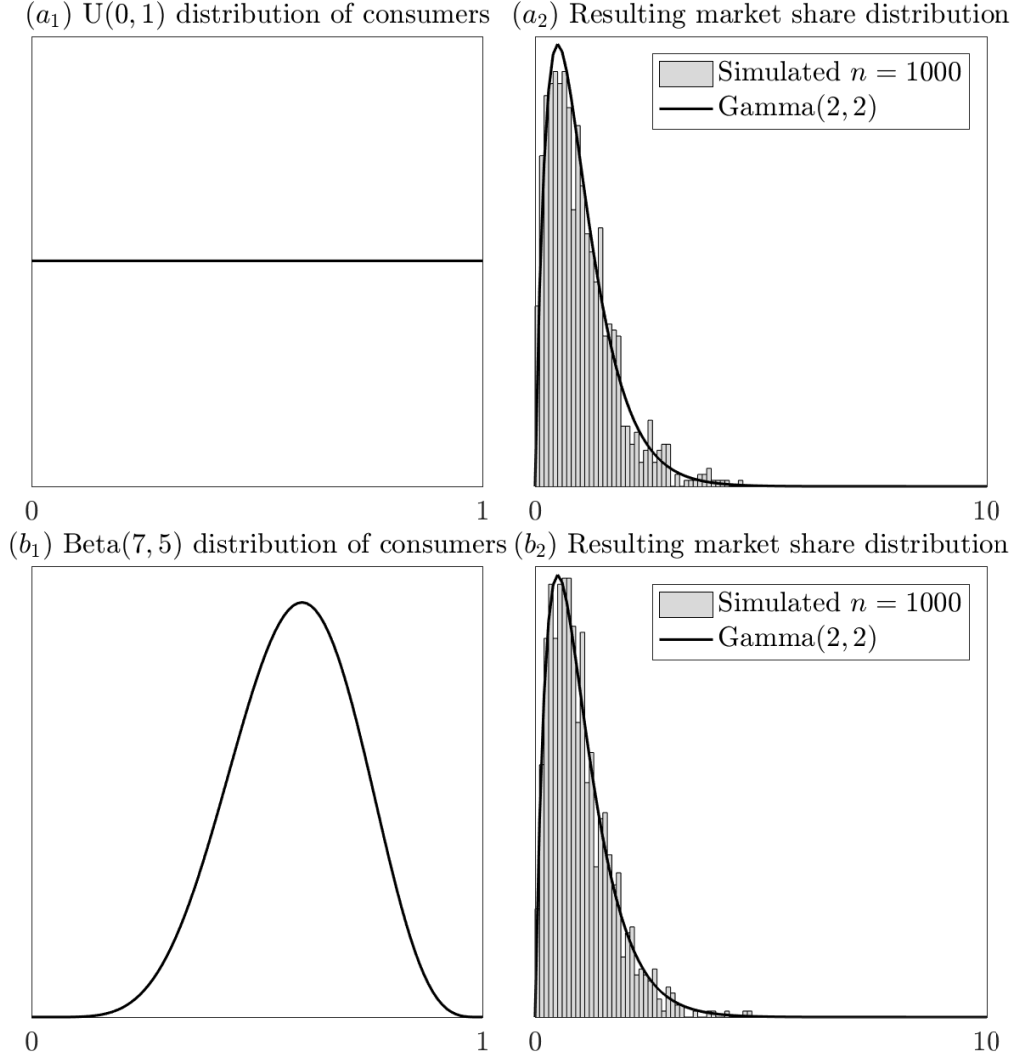
FIGURE 2. Simulation results.

of firms is large *regardless* of how the consumers are distributed on the unit interval. This key result therefore characterises the limiting market structure in the pure location game.

In political party competition (Xefteris, 2016; Ronayne, 2016), the unit interval represents the political spectrum (from left to right) and C represents the distribution of voters over that space. People vote for the party whose platform is closest to their position on the political spectrum and each political party chooses a platform (a location on the unit interval) to maximise its expected number of votes. Such a model is obviously formally equivalent to the pure location game described above. If the parties play the same strategy, then Theorem 1 tells us that the distribution of scaled party sizes (as measured by the fraction of vote shares)

will be Gamma$(2, 2)$ distributed when the number of parties is large *regardless* of how the voters are distributed on the political spectrum.[10]

A number of papers (Laster et al., 1999; Ottaviani and Sørensen, 2006; Ewerhart, 2015) have re-interpreted Hotelling's pure location game to model a contest between professional forecasters. A version of such a contest is given here: suppose that a public authority wants to know about some macroeconomic variable X but it does not know the distribution from which the macroeconomic variable is drawn. Forecasters learn that the variable is drawn from some distribution C on $[0, 1]$, and based on this, each forecaster sends a forecast to the public authority. The macroeconomic variable is then realised and the authority rewards the forecaster whose forecast was closest to the realisation. One can verify that the expected payoff of a forecaster who sends forecast x while all other forecasters choose their forecast according to the distribution L_n is given by (1). In other words, this contest game is strategically equivalent to Hotelling's pure location game. When a forecaster i chooses a forecast at x and the nearest other forecasts to the left and to the right of x are z and y respectively, then $C((x + y)/2) - C((x + z)/2)$ is the probability that i wins the contest. Now, suppose that the forecasters all play the same strategy and let P be a random variable representing the probability that an arbitrarily chosen forecaster wins the contest (before the realisation of the macroeconomic variable). Theorem 1 tells us that P is approximately Gamma$(2n, 2n)$ distributed when the number of forecasters is large *regardless* of how the macroeconomic variable is distributed.

I provide a more detailed discussion of Theorem 1 in section 6. For example, I explain what is behind the invariance of the limiting distribution, I explain the reason the limiting distribution of market shares is unequal even though the game is symmetric and the firms employ the same strategy, and I provide some intuition for why the limiting distribution is Gamma$(2, 2)$.

The proof of Theorem 1 builds on the theory of statistical spacings, which I present in section 4 below. This will provide the necessary background to develop intuition for the main result, and I use the tools of section 4 to present a heuristic proof of Theorem 1 in section 5; the formal proof of the result is in the appendix. I then use the intuition developed in sections 4 and 5 to provide the more detailed discussion of Theorem 1 in section 6.

4. Statistical spacings

Let us again consider a sample of n random variables $\mathcal{X} := \{X_1, ..., X_n\}$ in which each X_i is drawn independently from a parent distribution F on $[0, 1]$ and denote the ith order statistic of the sample by $X_{(i)}$. The ordered variables $X_{(1)}, ..., X_{(n)}$ induce $n - 1$ intervals on

---

[10]Theorem 1 is a limiting result and may therefore not apply in situations in which the number of competing parties is small.
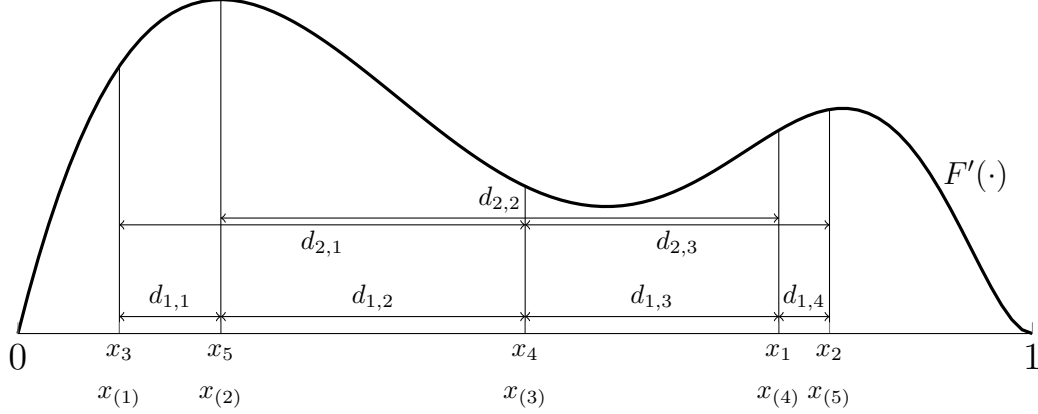
FIGURE 3. The above illustrates a possible realisation of a sample $\{X_1, ..., X_5\}$ of random variables each drawn from some distribution F. The realisations of $D_{1,i}$ and $D_{2,i}$ are denoted by $d_{1,i}$ and $d_{2,i}$ respectively.

$[0, 1]$ with lengths $D_{1,1}, ..., D_{1,n-1}$ where

$$(3) \qquad\qquad D_{1,i} := X_{(i+1)} - X_{(i)},$$

for $i \in \{1, ..., n-1\}$. The ordered variables $X_{(1)}, ..., X_{(n)}$ also induce $n-m$ intervals on $[0,1]$ with lengths $D_{m,1}, ..., D_{m,n-m}$ where

$$(4) \qquad\qquad D_{m,i} := X_{(i+m)} - X_{(i)},$$

for $i \in \{1, ..., n-m\}$. Each $D_{1,i}$ is a spacing of order one and each $D_{m,i}$ is a spacing of order $1 \leq m \ll n$ (David and Nagaraja, 2003). Statistical spacings are therefore random variables that are defined as the differences between order statistics. Spacings of order m are obviously a generalisation of spacings of order one, but it will be useful for purposes of exposition to treat spacings of order one on their own. Spacings of order one and two are illustrated in Figure 3.

In subsection 4.1 below, I present a result originally due to Weiss (1955) on the limiting empirical distribution of spacings of order one for an arbitrary parent distribution F. I extend the result to spacings of order m in subsection 4.2 and then use this extension for the case $m = 2$ to provide a heuristic proof for Theorem 1 in section 5.

4.1. **Spacings of order one.** Define $H_{1,n}(x) := \frac{1}{n-1} \sum_{i=1}^{n-1} \mathbf{1}(nD_{1,i} \leq x)$, where $D_{1,i}$ is given in (3). This is the empirical fraction of spacings of order one (scaled by n) whose length does not exceed some value $x \geq 0$. The following result is originally due to Weiss (1955), and it was later proved using a different method by Pyke (1965).

Proposition 2 (Weiss, 1955; Pyke, 1965). *If F is absolutely continuous then as $n \to \infty$,*

$$\sup_{x \geq 0} \left| H_{1,n}(x) - H_1(x) \right| \xrightarrow{p} 0,$$

where

$$H_1(x) := 1 - \int_{-\infty}^{\infty} \exp\{-xF'(z)\}dF(z).$$

One way to understand Proposition 2 is to consider the following thought experiment: break a stick of unit length n times with the location of each break-point drawn independently according to the distribution F on $[0,1]$. Then by Proposition 2, as n becomes large, the resulting distribution of (scaled) stick piece lengths is given by $H_1(\cdot)$. The image of "stick breaking", which was also employed by Holst (1980) in his study of spacings, might appear to be somewhat abstract but it will help to fix ideas and to highlight the important features and implications of Proposition 2. In particular, I use the thought experiment to make the related points (i)-(v) below. These will help in developing some insight into the underlying stochastic process and in drawing connections with standard models in economics and operations research.[11]

(i) The limiting distribution of stick piece lengths is positively skewed. Indeed, it is easy to verify that $H_1(x)$ is strictly increasing and strictly concave for any F, so a straightforward application of Jensen's inequality yields $H_1(\mathbb{E}[X]) > \mathbb{E}[H_1(X)] = 1/2$ for $X \sim H_1$. In other words, the distribution will be unequal with the majority of stick pieces being of smaller than average length *regardless* of the distribution F from which the break-points were drawn. To see why this is the case, notice that even when F is the uniform distribution, the independence of the draws implies that some break-points will *happen* to be drawn closer together than others, and this is what generates the inequality in the limiting distribution of stick piece lengths. In fact, one can show that for large n the density F' that maximises the fraction of pieces of greater than average length is the uniform: $F'(z) = 1$ for $z \in [0,1]$.[12] Indeed, any other density must have at least one local maximum but break-points are more likely to be drawn near a local maximum and will therefore be drawn close together, which generates more short pieces. Note that when the parent distribution F is the uniform distribution, the limiting

---

[11]Proposition 2 can for example be used to understand inter-arrival times in queues (Cooper, 1981). Imagine that the interval $[0,1]$ represents the length of a day and n people arrive uniformly throughout the day (so $F(z) = z$ for all $z \in [0,1]$). When n is large, the distribution of (scaled) inter-arrival times is given by $H_1(x) = 1 - \exp\{-x\}$, which is the cumulative distribution of an exponential random variable with mean one. A straightforward transformation shows that the distribution of (appropriately scaled) inter-arrival times over the period of one hour is exponential with mean $1/24$.

[12]When n is large, the average piece length (*scaled by n*) is $\mathbb{E}[X] = \int_0^\infty \Pr(X \geq x)dx = \int_0^\infty (1 - H_1(x))dx = 1$ for any F (and $X \sim H_1$). We want to choose the cumulative distribution function F with domain $[0,1]$ that maximises $(1 - H_1(1))$, i.e. the fraction of greater than average piece lengths. This is equivalent to choosing F' to maximise $\int_0^1 \exp\{-F'(z)\}F'(z)dz$. But note that for $y \geq 0$, $\exp\{-y\}y$ is maximised at $y = 1$. It follows that $\int_0^1 \exp\{-F'(z)\}F'(z)dz$ is maximised at $F'(z) = 1$ for all $z \in [0,1]$.

distribution of stick piece lengths is exponential with parameter one, i.e. $H_1'(x) = \exp\{-x\}$, which exhibits a significant degree of inequality.

(ii) The fact that the break-points are drawn independently from the same parent distribution is crucial for the inequality and the positive skewness that are exhibited by the limiting distribution. One can, for example, generate equal piece lengths if the break-points are drawn non-independently from the same non-degenerate distribution.[13] One can also generate a negatively skewed distribution, with a majority of stick pieces having greater than average length, if the break-points are drawn from different distributions.[14]

(iii) One should distinguish between procedures that generate break-points in an i.i.d. manner and procedures that generate piece lengths (i.e. intervals) in an i.i.d. manner. When break-points X_i are chosen in an i.i.d. manner, which is what happens in Proposition 2, the resulting piece lengths $D_{1,i}$ are *not* independent because the sum of the piece lengths is fixed. This phenomenon drives the shape of the limiting distribution and it is, incidentally, this induced non-independence of the spacings that complicates the proof of limit results relating to them. Contrast this with a procedure in which intervals lengths Y_i are drawn i.i.d. from some distribution J. Then the empirical distribution of the interval lengths, $\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(Y_i \leq x)$, converges straightforwardly to $J(x)$ for each x.[15]

(iv) Notice that the limiting distribution $H_1(x)$ in Proposition 2 depends on the parent distribution F, while the limiting distribution $G(x)$ in Theorem 1 does not depend on the limiting parent distribution $L = C$. But an invariance result does exist for spacings of order one. Recall that if X_i is drawn from a distribution F then, by the probability integral transform, $U_i := F(X_i)$ is a uniform random variable on $[0,1]$. The same is true for the order statistics. Namely, $U_{(i)} := F(X_{(i)})$ has the distribution of the ith order statistic from a uniform parent distribution on $[0,1]$. Now consider the spacings defined by

$$F(X_{(i+1)}) - F(X_{(i)}).$$

---

[13]For example, suppose n is even and draw X_1 and X_2 from F independently. Then for the ith draw, where $i > 2$, let $X_i = X_1$ when i is odd and $X_i = X_2$ when i is even. That is, the variables X_i for $i > 2$ have a particular correlation structure that depends entirely on the realisation of X_1 and of X_2 (but all variables have the same *marginal* distribution). Then, all of the resulting overlapping spacings will have length $|X_2 - X_1|$. *Physically*, if we ignore the end pieces, this breaking process would generate only one stick piece of length $|X_2 - X_1|$.

[14]For example, draw X_1 from a distribution F_1 that is degenerate at a point $p_1 = 0$ and draw X_2 from a distribution F_2 that is degenerate at a point p_2 that is strictly between 0 and $1/(n-1)$. Then draw each X_i for $i \in \{3,...,n\}$ from a distribution F_i that is degenerate at the point $p_i = p_2 + (i-2)(1-p_2)/(n-2)$, which equally divides the interval $[p_2, 1]$. The result will be $n-1$ stick pieces with the left-most piece having length $p_2 < 1/(n-1)$ while all other pieces have length $(1-p_2)/(n-2) > 1/(n-1)$.

[15]In the non-cooperative Nash demand game players demand piece lengths Y_i directly (i.e. they choose intervals) and each player receives their demanded piece length if the aggregate length of demanded pieces is less than one (i.e. the total stick length). When the demands are i.i.d. (which can occur in a mixed Nash equilibrium of the game, e.g. see Example 3 in Malueg, 2010) the empirical distribution of the piece lengths would need to be examined contingent on $\sum_{i=1}^{n} Y_i \leq 1$.

These spacings represent the integral of the density function F' between two consecutive break-points in the unit interval. But by the reasoning above they are also the first order differences between ordered random variables whose parent distribution is uniform on $[0, 1]$. By Proposition 2 it follows that the empirical distribution of these spacings (scaled by n) is given by $H_1(x)$ with $F(z) = z$ for all $z \in [0, 1]$. In other words, the empirical distribution of these spacings is exponential with parameter one. Remarkably, this is true regardless of the parent distribution F. A similar phenomenon is at play in obtaining the invariance of the gamma distribution in Theorem 1 since the market shares defined by (2) are essentially the integral of the density of consumers between consecutive mid-points of break-points in the unit interval and the break-points are chosen according to the density of consumers.

(v) A straightforward transformation of variables can make Proposition 2 applicable to new scenarios. For example, consider a situation in which gold panners (people who mine a riverbed for gold deposits using pans) place themselves independently at random along a river of length one (with source at zero) according to the distribution F on $[0, 1]$. Each panner searches for gold in the area between their location and the location of the next panner downstream from them. The value (or abundance) of gold at location x is given by some strictly positive value function $v(x)$. The value function is ex-ante unknown and the panners' location choice is non-strategic; it is determined by purely exogenous factors relating, for example, to the ease of access to various parts of the river bank. Let $V(x) := \int_0^x v(z)dz$. Then the value of gold collected by the ith panner can be written as

$$V(X_{(i+1)}) - V(X_{(i)}),$$

where each location X_i is drawn from the distribution F.

Remark 1. The limiting empirical distribution of the value of gold collected by the individual panners (scaled by n) is given by

$$(5) \qquad\qquad 1 - \int_{-\infty}^{\infty} \exp\left\{-x\frac{F'(z)}{v(z)}\right\} dF(z).$$

In other words, $F'(z)$ in the definition of $H_1(x)$ is just replaced by $F'(z)/v(z)$. A proof of this (that applies a transformation of variables) is given in the appendix and a simulation illustrating this remark is given in appendix B. ∎

The gold panner scenario illustrates the flexibility of the result: here, the value of the locations is completely independent of the manner in which the locations are chosen. Remark 1 tells us that the resulting limiting distribution is a type of exponential distribution modulated

by F and V. Note that when $V = F$ we recover the exponential distribution, just as was found in point (iv) above.[16]

Boppana et al. (2016) and Núñez and Scarsini (2017) generalise Hotelling competition on a line by studying spatial competition over a surface (e.g. a unit square), and the former set of authors also use the example of miners staking land claims to motivate their model. In these spatial games, each player chooses a location on the surface that maximises the Voronoi cell of their location and the players are restricted to choosing from a finite number of locations on the metric space. The model described here restricts the space to a line and is not strategic but the gold panners can locate at any point on the line and the value of locations along the river is not necessarily uniform.

4.2. **Spacings of order** m. I now considerably generalise Proposition 2 by extending the limiting result to spacings of order $1 \leq m \ll n$. Define $H_{m,n}(x) := \frac{1}{n-m} \sum_{i=1}^{n-m} \mathbf{1}(n D_{m,i} \leq x)$, where $D_{m,i}$ is given in (4). This is the empirical fraction of spacings of order m (scaled by n) whose length does not exceed some value $x \geq 0$.

Proposition 3. *If F is absolutely continuous then as $n \to \infty$,*

$$\sup_{x \geq 0} \left| H_{m,n}(x) - H_m(x) \right| \xrightarrow{p} 0,$$

where

$$H_m(x) := 1 - \int_{-\infty}^{\infty} \left(\sum_{k=0}^{m-1} \frac{(F'(z)x)^k}{k!} \right) \exp\{-F'(z)x\} dF(z).$$

This result will be key in allowing for the heuristic proof of Theorem 1 that is given in the next section. It will also be useful in exploring extensions of Hotelling's classic model, and I discuss this in section 6.

Remark 2. We can re-apply the transformation of variables that was previously done in Remark 1 to the case of spacings of order m. If each X_i is drawn independently from some distribution F and $v(x)$ is a strictly positive value function over $[0, 1]$ with $V(x) := \int_0^x v(z)dz$, then the limiting empirical distribution of the spacings $V(X_{(i+m)}) - V(X_{(i)})$ (scaled by n) is given by

$$1 - \int_{-\infty}^{\infty} \left(\sum_{k=0}^{m-1} \frac{(F'(z)x/v(z))^k}{k!} \right) \exp\{-F'(z)x/v(z)\} dF(z).$$

---

[16]The gold panners scenario bears some formal resemblance to cake-cutting games. A cake is represented by the unit interval and each player i has a value function v_i over the cake. The goal for a designer is to cut the cake in a manner that satisfies desirable criteria such as envy-freeness. Within the framework considered here, the designer would choose the function F according to which the cuts would be drawn at random. But it is not clear that this approach is best suited to answer the questions typically posed in the literature and, as noted by Chen et al. (2013), few papers have studied randomised designs.

In other words, each occurrence of $F'(z)$ in the definition of $H_m(x)$ is simply replaced by $F'(z)/v(z)$. I omit the proof since it follows exactly the same steps as the proof of Remark 1. Furthermore, taking a derivative of the equation above with respect to x yields

$$(6) \qquad \int_{-\infty}^{\infty} \frac{(F'(z)/v(z))^m}{(m-1)!} x^{m-1} \exp\{-(F'(z)/v(z))x\}dF(z),$$

which is the density at x of the limiting empirical distribution of the scaled spacings $V(X_{(i+m)}) - V(X_{(i)})$. ∎

Proposition 3 and Remark 2 have an important connection with the gamma family of distributions. The gamma distribution is parametrised by a "shape" parameter $a > 0$ and a "rate" parameter $b > 0$; and note that b^{-1} is often referred to as a "scale" parameter. At any $x > 0$ the density function of a Gamma(a, b) random variable is given by

$$(7) \qquad \frac{b^a}{(a-1)!} x^{a-1} \exp\{-bx\}.$$

The exponential distribution is a special case in which $a = 1$.

Comparing (7) and (6), one can see that the limiting empirical distribution of the scaled spacings $V(X_{(i+m)}) - V(X_{(i)})$ is clearly the density of a gamma-like distribution modulated by F and V. In fact, whenever $V = F$ the distribution is exactly a Gamma$(m, 1)$.

Remark 3. Consider the spacings

$$\frac{1}{c}[U_{(i+m)} - U_{(i)}],$$

where $c > 0$ is some constant and each $U_{(i)}$ is an order statistic whose parent distribution is a uniform on $[0, 1]$. Then using Remark 2 it suffices to set $V(x) = x/c$ and $F(z) = z$ for all $z \in [0, 1]$ to see that the limiting empirical distribution of these spacings (scaled by n) will be characterised by the density (??) with $b = c$ and $a = m$. This is precisely the density of a Gamma(m, c) random variable. In other words, the spacing order, m, becomes the shape parameter of the gamma distribution while the constant, c, becomes the rate (and its reciprocal becomes the scale) parameter. ∎

5. Heuristic proof of Theorem 1

I present a heuristic proof of Theorem 1 in this section. The market shares S_i given in (2) are not quite spacings as they are defined in section 4. However, as I show below, they can be re-expressed as spacings of order two when n is large.

The market shares defined in (2) can be written as

$$S_i := C\left(\frac{X_{(i+2)} + X_{(i+1)}}{2}\right) - C\left(\frac{X_{(i+1)} + X_{(i)}}{2}\right)$$

$$\approx \frac{1}{2}\left[C(X_{(i+2)}) + C(X_{(i+1)}) - C(X_{(i+1)}) - C(X_{(i)})\right]$$

$$= \frac{1}{2}\left[C(L_n^{-1}(U_{(i+2)})) - C(L_n^{-1}(U_{(i)}))\right]$$

$$\approx \frac{1}{2}\left[U_{(i+2)} - U_{(i)}\right].$$

The first line is simply the definition of S_i. The second line follows from the fact that when n becomes large, the points $(X_{(i)} + X_{(i+1)})/2$ and $(X_{(i+1)} + X_{(i+2)})/2$, at which the function C is being evaluated, become very close together since the locations of the many firms become more tightly packed on the interval. The function C is therefore approximately linear between the two points. Once C is pulled into the expression (by its local linearity) the term $C(X_{(i+1)})$ can be cancelled out. Since each X_i is drawn from the distribution L_n, the probability integral transform gives us the relationship $U_{(i)} = L_n(X_{(i)})$, and this is used in the third line. Finally, the last line follows from Osborne and Pitchik's (1986) result (Proposition 1); namely that $L_n \to L = C$ for large n.

The derivation above shows that $S_i \approx (U_{(i+2)} - U_{(i)})/2$ when n is large. Remark 3 then immediately gives us that the limiting empirical distribution of the spacings $(U_{(i+2)} - U_{(i)})/2$ (scaled by n) is a Gamma(2, 2) distribution. Of course, the derivation here is only heuristic but the proof in the appendix provides a formal argument.

6. DISCUSSION

Many features of Theorem 1 can be explained by applying the insights derived in sections 4 and 5. For example, the fact that the limiting distribution of market shares exhibits some inequality follows from the discussion in point (i) of section 4. The fact that the limiting distribution belongs to the gamma family follows from the result that the limiting empirical distribution of scaled spacings is closely related to this family (e.g. see Proposition 3).[17] In

---

[17]Different stochastic processes can lead to distributions belonging to other parametric families. For example, in the context of city growth, if all cities grow at the same average rate regardless of their size then the resulting city size distribution is either Pareto or log-normal (Gibrat, 1931; Gabaix, 2009). Indeed, if city i grows according to $\text{size}_{i,t+1} = \text{growth\_rate}_{i,t} \times \text{size}_{i,t}$, where the growth rates are i.i.d. random variables, then for large t the resulting city size distribution is log-normal. If a small but positive error is added to the auto-regressive process then we have a Kesten process and the resulting distribution is Pareto (Gabaix, 1999). The gamma parametric family is qualitatively different from the Pareto or log-normal families because gamma distributions have thinner right tails. The literature does contain examples of stochastic processes that generate distributions belonging to the gamma family. For example, Sutton (1998, Chapter 10.3) proposes a model in which an opportunity arises in each period that can be taken up by only one firm in that period. The limiting firm size distribution is exponential, which is a special case of the gamma family (and a firm's size is measured by the number of opportunities that it has accumulated over time). Angle (1993) proposes

fact, the limiting empirical distribution of scaled market shares belongs to a specific member of this family: it is Gamma$(2, 2)$ *regardless of how the consumers are distributed.* This invariance result follows from the fact that, in the limit, the location strategy of each firm mimics the distribution of consumers (and therefore, as shown in section 5, the market share of any firm can be expressed as a spacing of order two of variables whose parent distribution is uniform). If the locations of the firms were drawn according to a distribution that is different from the distribution of consumers then the limiting distribution of market shares would be modulated by the firms' location strategy as well as by the distribution of consumers (e.g. see Remark 2).

Of course, the invariance result relies on a number of assumptions which, if relaxed, might lead to different limiting market structures. In the original formulation of the location game given in section 2, each firm i essentially sells one unit of its good to each consumer who is located between the midpoint between i and i's nearest neighbours to the left and to the right (and each consumer therefore consumes the good of only one firm). Consider the following variant: each firm i sells one unit of its good to each consumer who is located between the midpoint between i and i's *second* nearest neighbours to the left and to the right. Consumers located between two adjacent firms might purchase from *each* of the two firms. Competitive pressure on each firm is reduced in this variant of the game because the markets of the firms overlap.[18] Demand is therefore less rival across firms. In this case, the payoff of firm $i + 2$ can be written as

$$C\left(\frac{X_{(i+4)} + X_{(i+2)}}{2}\right) - C\left(\frac{X_{(i+2)} + X_{(i)}}{2}\right).$$

Compared to the shares of the original game, which are given in (2), the payoff here is simply the integral of the distribution of consumers evaluated over a larger band of the unit interval.

Remark 4. Suppose that the distribution of consumers is uniform and that the firms play a symmetric equilibrium in which they choose their locations uniformly at random on the unit interval (this is indeed an equilibrium of the limiting game and a proof is given in appendix).[19] Then, it is straightforward to see that the payoffs can be written as $(U_{(i+4)} - U_{(i)})/2$. By Remark 3 the limiting empirical distribution of scaled payoffs is a Gamma$(4, 2)$ distribution. A simulation illustrating this remark is given in appendix B. ∎

---

a stochastic process in which whenever two people interact, one is chosen at random to gain a fraction of the other's wealth (and the loser retains only the remaining fraction). The resulting distribution of wealth belongs to the gamma family.

[18]Williams and Senior (1977) and De Palma et al. (1985) consider an extension that is similar in spirit.

[19]This remark merely presents *an* application of the methods developed in this paper to a variant of the pure location game. I therefore consider only the simple case in which the distribution of consumers is uniform because deriving the limiting symmetric equilibrium for this case is straightforward. One would need a result analogous to Proposition 1 adapted to this variant of the game in order to consider limiting equilibria under a non-uniform distribution of consumers.

Now note that if $X \sim \text{Gamma}(m, c)$ and $X' \sim \text{Gamma}(m', c)$ with $m < m'$ then X exhibits more inequality than X' in the Lorenz sense, i.e. the Lorenz curve associated with X lies everywhere below the Lorenz curve associated with X' (see Kleiber and Kotz, 2003, pp. 24, 164). Remark 4 therefore tells us that in the variant of the game in which competitive pressure is reduced the distribution of payoffs is less unequal.

There are many avenues for future work. First, I have assumed that the firms play a symmetric equilibrium. Arguably, such equilibria are focal in the symmetric game, especially when the number of firms is large. That said, one could examine what the market structure would be in situations in which the firms play different strategies from each other. Second, I have assumed that the firms choose their locations simultaneously and independently. Following point (ii) in section 4, the limiting distribution may be very different if correlated strategies were allowed or if firms chose their locations sequentially – which would imply that their choices can be drawn from different distributions (e.g. see Prescott and Visscher, 1977, or Loertscher and Muehlheusser, 2011). Third, the firms can locate at any point on the unit interval. Núñez and Scarsini (2017) consider a variant of the pure location game in which firms can locate on a finite number of locations over a surface.[20] Extending the results of the paper to a more general space would be fruitful, but it is likely to be technically challenging: points on an interval are completely ordered, so one can rely on the existing theory of order statistics (and their corresponding spacings) to derive the results of this paper. No such ordering of points exists on a two dimensional surface so new methods may need to be developed to derive limiting results. Fourth, the results presented in this paper apply, strictly speaking, only in the limit as the number of firms goes to infinity. Currently, no tractable characterisation exists of all symmetric equilibria for finite n and an arbitrary distribution of consumers. But if such a characterisation were found, one could then derive the market structure in Hotelling's pure location game for a finite number of firms. Fifth, I have only considered Hotelling's *pure* location game, which abstracts from the firms' pricing decision. Examining the market structure in a game that includes a pricing decision is a promising area of future research.

---

[20]If firms could choose from, say, only two globally fixed locations on the unit interval then the resulting limiting distribution of market shares would not belong to the gamma family since the distribution would be concentrated on only two points. Indeed, if the firms could locate only at points x or y where $0 < x < y < 1$ then the only possible resulting market shares are $(x + y)/2\mu_x$ and $1 - (x + y)/2\mu_y$ where μ_x is the number of firms located at x and μ_y is the number of firms located at y.

APPENDIX A. PROOFS

Proof of Remark 1. Let $Y_{(i)} := V(X_{(i)})$, where each X_i is drawn according to F. Then by a transformation of variables, each Y_i is drawn from a distribution F_Y satisfying

$$F_Y'(y) = F'(V^{-1}(y))\frac{\partial V^{-1}(y)}{\partial y} = \frac{F'(x)}{V'(x)} = \frac{F'(x)}{v(x)},$$

where $x = V^{-1}(y)$, and so $v(x)dx = dy$. Substituting into $H_1(\cdot)$ yields (5). $\qquad\square$

I now present the key steps of Pyke's (1965) proof of Proposition 2. For the reader interested in the technical aspects of this paper, reading this proof will be useful because I follow broadly the same steps to prove all the other limiting results.

Proof of Proposition 2. Since $H_1(\cdot)$ is continuous, monotonic, and bounded, the uniformity of the convergence that is given in the statement of the result will follow from pointwise convergence in probability (see Weiss, 1955). To prove the latter, we must show that for any $x \geq 0$ and any $\epsilon > 0$, $\lim_{n\to\infty}\Pr(|H_{1,n}(x) - H_1(x)| > \epsilon) = 0$. And, by Chebyshev's inequality, it suffices to show that $\mathbb{E}[H_{1,n}(x)] \to H_1(x)$ and $\mathrm{Var}[H_{1,n}(x)] \to 0$ as $n \to \infty$ for all $x \geq 0$.

The remaining difficulty is the evaluation of $\mathbb{E}[H_{1,n}(x)]$ and $\mathrm{Var}[H_{1,n}(x)]$: the variables $D_{1,i}$ and $D_{1,j}$ when $i \neq j$ are not independent and have different distributions, so taking the expectation or variance directly is not tractable. However, Pyke (1965) proposed an ingenious transformation of variables that allows us to proceed further. For each $X_i \in \mathcal{X}$ define the random variable

(8) $\qquad I_{1,i}(x,n) := \begin{cases} 1 & \text{if } no \text{ random variable in } \mathcal{X} \text{ falls in } (X_i, X_i + x/n] \\ 0 & \text{otherwise} \end{cases}$.

The variables $I_{1,i}(x,n)$ for $i \in \{1,...,n\}$ are not independent of each other but they are exchangeable. The key step to the proof is to see that

(9) $$H_{1,n}(x) = \frac{n}{n-1} - \frac{1}{n-1}\sum_{i=1}^n I_{1,i}(x,n).$$

To see why this is true, order the variables in $\mathcal{X}$ and denote them as $X_{(1)},...,X_{(n)}$. For each $j \in \{1,...,n\}$ the realisation of $X_{(j)}$ is the realisation of some variable X_i where the indices are related by $i = \sigma(j)$ and σ is a permutation on $\{1,...,n\}$. So if the first spacing $D_{1,1} = X_{(2)} - X_{(1)}$, say, is less than x/n it is because some variable in $\mathcal{X}$ has fallen between $X_{(1)}$ and $X_{(1)}+x/n$, but the realisation of $X_{(1)}$ is simply the realisation of some variable $X_{\sigma(1)}$. In other words, the statement $D_{1,1} \leq x/n$ if and only if $I_{1,\sigma(1)}(x,n) = 0$ will hold for the realisations of the random variables under the ex-post permutation σ. This relationship holds
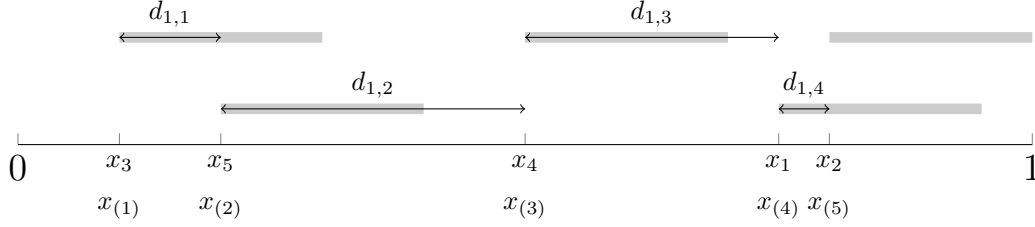
FIGURE 4. The grey bands represent the distance x/n for some $x > 0$, which can be compared with the lengths of the spacings of order one. For example, $d_{1,1} < x/n$ but $d_{1,2} > x/n$. The realisation of $\sum_{i=1}^{n-1} \mathbf{1}(nD_{1,i} \leq x)$ is clearly 2. Notice that, here, σ is the map $(1,2,3,4,5) \mapsto (3,5,4,1,2)$. The realisation of the indicator $I_{1,3}(x,n)$ is 0 since x_5 appears between x_3 and $x_3 + x/n$. The realisations of the indicators $I_{1,1}(x,n), ..., I_{1,n}(x,n)$ are, respectively, $0, 1, 0, 1, 1$. The realisation of $\sum_{i=1}^{n}(1 - I_{1,i}(x,n))$ is 2.

for $D_{1,1}, ..., D_{1,n-1}$. The realisation of the indicator $I_{1,\sigma(n)}(x,n)$ is always equal to one.[21] It follows that $\sum_{i=1}^{n-1} \mathbf{1}(nD_{1,i} \leq x) = \sum_{i=1}^{n}(1 - I_{1,i}(x,n))$, and dividing by $n-1$ yields equation (9). An illustration of this relationship is shown in Figure 4.

We can now derive the expectation $\mathbb{E}[H_{1,n}(x)]$ and its limit. Note that for each i,

$$\mathbb{E}[I_{1,i}(x,n)] = \Pr(\text{no random variable in } \mathcal{X} \text{ falls in } (X_i, X_i + x/n])$$

(10)
$$= \int_{-\infty}^{\infty} [1 - (F(z + x/n) - F(z))]^{n-1} dF(z).$$

Conditional on $X_i = z$ the term in square brackets to the $n-1$ is the probability that the $n-1$ variables X_k for $k \neq i$ fall outside the interval $(z, z + x/n]$. Taking the limit of (10) as $n \to \infty$ we obtain

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} [1 - (F(z + x/n) - F(z))]^{n-1} dF(z) = \int_{-\infty}^{\infty} \exp\{-F'(z)x\} dF(z).$$

Indeed, we can pull the limit inside the integral and the integral converges by the dominated convergence theorem. Somewhat informally, the limit of the term inside the integral can be obtained from the fact that for large n, $F(z + x/n) - F(z) \approx F'(z)x/n$, and that $\exp\{-F'(z)x\} = \lim_{n \to \infty} (1 - F'(z)x/n)^{n-1}$. Using (9), it is then easy to see that

$$\lim_{n \to \infty} \mathbb{E}[H_{1,n}(x)] = H_1(x).$$

To complete the proof, we still need to show that $\lim_{n \to \infty} \mathrm{Var}[H_{1,n}(x)] = 0$. Firstly note that $\mathbb{E}[I_{1,i}(x,n)^2] = \mathbb{E}[I_{1,i}(x,n)]$, and for $j \neq i$, the term $\mathbb{E}[I_{1,i}(x,n)I_{1,j}(x,n)]$ is equal to the probability that no variable in the sample $\mathcal{X}$ falls in $(X_i, X_i + x/n] \cup (X_j, X_j + x/n]$. This,

---

[21]For some k we must have that $k = \sigma(n)$. But then no random variable in the sample can have fallen above the realisation of X_k, so the indicator corresponding to this variable must have a realised value of one.

in turn, is equal to

$$(11) \quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[1-\{(F(z+x/n)-F(z))+(F(y+x/n)-F(y))-\zeta(x,y,z,n)\}]^{n-2}dF(z)dF(y).$$

Conditional on $X_i = y$ and $X_j = z$, the term in curly brackets is the probability that X_k falls in the interval $(y, y + x/n] \cup (z, z + x/n]$, where $k \neq i, j$. In particular, $F(z + x/n) - F(z)$ is the probability of falling in the interval $(z, z + x/n]$, $F(y + x/n) - F(y)$ is the probability of falling in the interval $(y, y + x/n]$, and $\zeta(x, y, z, n)$ is the probability of falling in their intersection, and it is defined as

$$(12) \qquad \zeta(x,y,z,n) := \begin{cases} 0 & \text{if } z \in (-\infty, y - x/n) \\ F(z + x/n) - F(y) & \text{if } z \in [y - x/n, y) \\ F(y + x/n) - F(z) & \text{if } z \in [y, y + x/n) \\ 0 & \text{if } z \in [y + x/n, \infty) \end{cases}.$$

The term in square brackets to the $n - 2$ in equation (11) is the probability that all $n - 2$ variables X_k fall outside the interval. Notice that the intervals over which $\zeta(x, y, z, n)$ may be non-zero vanish as $n \to \infty$.

Let $\alpha_1(x, n)$ be defined as the expression given in (10) and let $\alpha_2(x, n)$ be defined as the expression given in (11). Following the same reasoning that was used to obtain the limit of $\alpha_1(x, n)$, we have that[22]

$$\lim_{n\to\infty} \alpha_2(x,n) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\{-(F'(y)+F'(z))x\}dF(z)dF(y)$$

$$= \left(\int_{-\infty}^{\infty} \exp\{-F'(z)x\}dF(z)\right)^2.$$

We can now use this result to show that the variance of $H_{1,n}(x)$ goes to zero in the limit as $n \to \infty$. Obviously, $\text{Var}[H_{1,n}(x)] = \mathbb{E}[H_{1,n}(x)^2] - \mathbb{E}[H_{1,n}(x)]^2$. But $\mathbb{E}[H_{1,n}(x)^2]$ is equal to

$$\mathbb{E}\left[\frac{n^2}{(n-1)^2} - \frac{2n}{(n-1)^2}\sum_{i=1}^{n} I_i^1(x,n) + \frac{1}{(n-1)^2}\sum_{i=1}^{n}\sum_{j=1}^{n} I_i^1(x,n)I_j^1(x,n)\right].$$

The expression above becomes

$$(13) \qquad \frac{n^2}{(n-1)^2} - \frac{2n}{(n-1)^2}n\alpha_1(x,n) + \frac{n}{(n-1)^2}\alpha_1(x,n) + \frac{n(n-1)}{(n-1)^2}\alpha_2(x,n).$$

Taking the limit of (13) as $n \to \infty$, we are left with

$$\lim_{n\to\infty} \mathbb{E}[H_{1,n}(x)^2] = 1 - 2\lim_{n\to\infty} \alpha_1(x,n) + \lim_{n\to\infty} \alpha_2(x,n) = H_1(x)^2 = \lim_{n\to\infty} \mathbb{E}[H_{1,n}(x)]^2.$$

---

[22]For intuition, when n is large, $I_{1,i}(x, n)$ and $I_{1,j}(x, n)$ essentially become independent since X_i and X_j are drawn independently. Therefore $\mathbb{E}[I_{1,i}(x, n)I_{1,j}(x, n)] \approx \mathbb{E}[I_{1,i}(x, n)]\mathbb{E}[I_{1,j}(x, n)] = \mathbb{E}[I_{1,i}(x, n)]^2$ for large n.

This completes the proof. □

Notice that in Proposition 2, each X_i is drawn from some parent distribution F directly rather than being drawn from some distribution F_n that tends to F as n becomes large. This contrasts with Theorem 1, where each X_i has a parent distribution L_n that may be different from C but that effectively tends to C as n become large. However, as Remark 5 below shows, Proposition 2 can be easily modified to accommodate for a parent distribution that varies with n.

Remark 5. Instead of the X_i being drawn from an absolutely continuous parent distribution F, suppose that they are drawn from a parent distribution $F_n \in \mathcal{C}^2$ such that $F_n \to F$, $F'_n \to F'$, $F''_n \to F''$ uniformly for some $F \in \mathcal{C}^2$. Then, the limiting distribution of the resulting spacings of order one would still be $H_1(\cdot)$. Indeed, one can repeat all the steps of the proof of Proposition 2 above with the only difference that now the parent distribution has an "n" subscript. Taking the limit needs somewhat more work but ultimately, as I show below, the limiting result remains unchanged. ■

Proof of Remark 5. In this case equation (10) becomes

$$(14) \qquad \mathbb{E}[I_{1,i}(x,n)] = \int_{-\infty}^{\infty} [1 - P_n(x,z)]^{n-1} dF_n(z),$$

where $P_n(x,z) := F_n(z + x/n) - F_n(z)$. The only difference is that F in (10) is now F_n in (14). Now notice that

$$\lim_{n\to\infty} [1 - P_n(x,z)]^{n-1} = \exp\left\{ \lim_{n\to\infty} (n-1)\ln(1 - P_n(x,z)) \right\}.$$

Using the inequality $-y(1-y)^{-1} \leq \ln(1-y) \leq -y$ for $y \in [0,1)$ one obtains

$$-(n-1)P_n(x,z)(1 - P_n(x,z))^{-1} \leq (n-1)\ln(1 - P_n(x,z)) \leq -(n-1)P_n(x,z).$$

By Taylor's theorem, $F_n(z + x/n) = F_n(z) + (x/n)F'_n(z) + (1/2)(x/n)^2 F''_n(t_n)$, where the last term is the Lagrange form of the remainder and t_n is a number between z and $z + x/n$. Since $F''_n \to F''$ uniformly and $t_n \to z$ as $n \to \infty$ we have that $F''_n(t_n) \to F''(z)$ (e.g. see Rudin, 1976). It follows that $-(n-1)P_n(x,z) \to -xF'(z)$ and $-(n-1)P_n(x,z)(1 - P_n(x,z))^{-1} \to -xF'(z)$ as $n \to \infty$, and so $\lim_{n\to\infty}(n-1)\ln(1 - P_n(x,z)) = -xF'(z)$. In other words, (14) has the same limit as (10), so this takes care of the expectation of the empirical distribution. The convergence of its variance at each point to zero can be dealt with in a similar manner. □

Proof of Proposition 3. Just as was the case in the proof of Proposition 2, since $H_m(x)$ is continuous, monotonic, and bounded, it suffices to show that $\mathbb{E}[H_{m,n}(x)] \to H_m(x)$ and

$\text{Var}[H_{m,n}(x)] \to 0$ as $n \to \infty$ for all x. Now, for each $X_i \in \mathcal{X}$ define the random variable

$$I_{m,i}(x,n) := \begin{cases} 1 & \text{if } at\ most\ m-1 \text{ random variables in } \mathcal{X} \text{ fall in } (X_i, X_i + x/n] \\ 0 & \text{otherwise} \end{cases}.$$

The variables $I_{m,i}(x,n)$ are similar to the variables $I_{1,i}(x,n)$ defined in (8) in the proof of Proposition 2, with the exception of "no" in the definition of $I_{1,i}(x,n)$ being replaced with "at most $m-1$" in the definition of $I_{m,i}(x,n)$. The variables $I_{m,i}(x,n)$ now allow us to write[23]

$$H_{m,n}(x) = \frac{n}{n-m} - \frac{1}{n-m} \sum_{i=1}^{n} I_{m,i}(x,n).$$

The rest of the proof proceeds in the same way as the proof of Proposition 2. To start, note that $\mathbb{E}[I_{m,i}(x,n)]$ is equal to

$$\Pr(\text{at most } m-1 \text{ random variables in } \mathcal{X} \text{ fall in } (X_i, X_i + x/n])$$

$$(15) \quad = \int_{-\infty}^{\infty} \sum_{k=0}^{m-1} \binom{n-k}{k} \{F(z+x/n) - F(z)\}^k [1 - (F(z+x/n) - F(z))]^{n-k-1} dF(z).$$

Conditional on $X_i = z$ the term in the curly brackets to the k is the probability that k of the remaining $n-1$ random variables fall in the interval $(z, z+x/n]$ while the term in square brackets to the $n-k-1$ is the probability that the other $n-k-1$ random variables fall outside the interval. The probabilities are multiplied by the number of ways that each event can occur and aggregated over all k from 0 to $m-1$.

Define $\beta_1(x,n)$ as the expression given in (15). Since $n-k$ choose k is equal to $1/k!$ multiplied by a term of order n^k, we have that

$$\lim_{n \to \infty} \beta_1(x,n) = \int_{-\infty}^{\infty} \left(\sum_{k=0}^{m-1} \frac{(F'(z)x)^k}{k!} \right) \exp\{-F'(z)x\} dF(z).$$

This time, expanding $\mathbb{E}[H_{m,n}(x)]$ and taking its limit one obtains

$$\lim_{n \to \infty} \mathbb{E}[H_{m,n}(x)] = 1 - \lim_{n \to \infty} \beta_1(x,n) = H_m(x).$$

Having found the limit of the expectation, let us now turn to the variance. Notice that $\mathbb{E}[I_{m,i}(x,n)I_{m,j}(x,n)]$ is the probability that at most $m-1$ random variables fall in $(X_i, X_i + x/n]$ and at most $m-1$ random variables fall in $(X_j, X_j + x/n]$. This probability, which I

---

[23]This time, returning to the ex-post permutation σ discussed in the proof of Proposition 2, $D_{m,i} \leq x/n$ if and only if $I_{m,\sigma(j)}(x,n) = 0$ for $j \in \{1, ..., n-m\}$, but the realisations of $I_{m,\sigma(n-m+1)}(x,n), ..., I_{m,\sigma(n)}(x,n)$ are always equal to one. Therefore $\sum_{i=1}^{n-m} \mathbf{1}(nD_{m,i} \leq x) = \sum_{i=1}^{n} (1 - I_{m,i}(x,n))$.

denote by $\beta_2(x,n)$, is equal to

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sum_{a=0}^{m-1}\sum_{b=0}^{m-1}\binom{n-2}{a+b}\times$$

$$\binom{a+b}{a}\left[\{Q_n(y)-\zeta\}^a\{Q_n(z)\}^b+\sum_{t=0}^{a-1}\binom{a}{t}\{Q_n(y)-\zeta\}^t\zeta^{a-t}\{Q_n(z)-\zeta\}^{b-(a-t)}\right]\times$$

$$[1-\{Q_n(z)+Q_n(y)-\zeta\}]^{n-2-(a+b)}dF(z)dF(y).$$

where ζ is being used as short-hand for $\zeta(x,y,z,n)$ as it is defined in (12) and $Q_n(q) :=$ $F(q+x/n)-F(q)$. One can understand this equation as follows: fix $X_i = y$ and $X_j = z$ and consider what happens to the $n-2$ remaining variables. Choose $a+b$ of the $n-2$ variables to fall outside $(y, y+x/n] \cup (z, z+x/n]$ (this is the term in the small square brackets to the $n-2-(a+b)$). For the $a+b$ variables that fall inside $(y, y+x/n] \cup (z, z+x/n]$, the term in the large square brackets is the probability that exactly a of them fall in $(y, y+x/n]$ and b of them fall in $(z, z+x/n]$, accounting for the fact that the intervals may intersect. The terms are multiplied by the number of ways one can choose $a+b$ from $n-2$ and a from $a+b$. These are them summed over a and b from 0 to $m-1$.

Note that $n-2$ choose $a+b$ multiplied by $a+b$ choose a is equal to $1/(a!b!)$ times a term of order n^{a+b}. Furthermore, the term ζ vanishes as n becomes large (since the intersection of the intervals goes to zero). Taking the limit as $n \to \infty$ we therefore obtain that $\lim_{n\to\infty}\beta_2(x,n)$ is equal to

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sum_{a=0}^{m-1}\sum_{b=0}^{m-1}\frac{(F'(y)x)^a}{a!}\frac{(F'(z)x)^b}{b!}\exp\{-(F'(y)+F'(z))x\}dF(z)dF(y)$$

$$=\left(\int_{-\infty}^{\infty}\left(\sum_{k=0}^{m-1}\frac{(F'(z)x)^k}{k!}\right)\exp\{-F'(z)x\}dF(z)\right)^2.$$

We can now use this result to show that the variance of $H_{m,n}(x)$ goes to zero in the limit as $n \to \infty$. Obviously, $\text{Var}[H_{m,n}(x)] = \mathbb{E}[H_{m,n}(x)^2] - \mathbb{E}[H_{m,n}(x)]^2$. But expanding $\mathbb{E}[H_{m,n}(x)^2]$ we obtain that it is equal to

(16) $$\frac{n^2}{(n-m)^2} - \frac{2n}{(n-m)^2}n\beta_1(x,n) + \frac{n}{(n-m)^2}\beta_1(x,n) + \frac{n(n-1)}{(n-m)^2}\beta_2(x,n).$$

If we take the limit of (16) as $n \to \infty$, we are left with

$$\lim_{n\to\infty}\mathbb{E}[H_{m,n}(x)^2] = 1 - 2\lim_{n\to\infty}\beta_1(x,n) + \lim_{n\to\infty}\beta_2(x,n) = H_m(x)^2 = \lim_{n\to\infty}\mathbb{E}[H_{m,n}(x)]^2.$$

This completes the proof. $\qquad\square$

24

Proof of Theorem 1. Partition the unit interval into $A(\delta)$ subintervals (a_i, a_{i+1}) for $i \in \{0, ..., A(\delta) - 1\}$, with $a_0 = 0$ and $a_{A(\delta)} = 1$, such that over any given union of adjacent subintervals $(a_i, a_{i+1}) \cup (a_{i+1}, a_{i+2})$ and any $x, y \in (a_i, a_{i+2})$ the continuous density $C'(\cdot)$ satisfies $|C'(x) - C'(y)| < \delta$. For any particular subinterval (a_i, a_{i+1}) denote by $\mathcal{X}(i)$ the set of random variables from the sample $\mathcal{X}$ that fall into this interval and denote by $X^i_{(1)}, ..., X^i_{(|\mathcal{X}(i)|)}$ the ordered random variables that are in $\mathcal{X}(i)$. For $j \in \{1, ..., |\mathcal{X}(i)| - 2\}$ let us define the spacings

$$S^i_j := C\left(\frac{X^i_{(j+2)} + X^i_{(j+1)}}{2}\right) - C\left(\frac{X^i_{(j+1)} + X^i_{(j)}}{2}\right)$$

$$= \frac{X^i_{(j+2)} - X^i_{(j)}}{2} C'(Y^i_j)$$

$$= \frac{X^i_{(j+2)} - X^i_{(j)}}{2} (C'(X^i_{(j)}) + Z^i_{(j)}).$$

The second line follows from the mean value theorem for some $Y^i_j \in [(X^i_{(j)} + X^i_{(j+1)})/2, (X^i_{(j+1)} + X^i_{(j+2)})/2]$ and the third line follows from setting $C'(X^i_{(j)}) + Z^i_{(j)} = C'(Y^i_j)$. The spacings defined above are determined by cuts that fall strictly within (a_i, a_{i+1}). Now notice that if (a_i, a_{i+1}) is not the right-most interval then there are the two spacings which we may define as $S^i_{|\mathcal{X}(i)|-1} := (X^{i+1}_{(1)} - X^i_{(|\mathcal{X}(i)|-1)})(C'(X^i_{(|\mathcal{X}(i)|-1)}) + Z^i_{(|\mathcal{X}(i)|-1)})/2$ and $S^i_{|\mathcal{X}(i)|} := (X^{i+1}_{(2)} - X^i_{(|\mathcal{X}(i)|)})(C'(X^i_{(|\mathcal{X}(i)|)}) + Z^i_{(|\mathcal{X}(i)|)})/2$ that are determined by cuts that straddle (a_i, a_{i+1}) and (a_{i+1}, a_{i+2}). So for each $i < A(\delta) - 1$ and each $j \in \{1, ..., |\mathcal{X}(i)|\}$ we can think of $Z^i_{(j)}$ as being equal to some function $\Delta^i(X^i_{(j)}, \mathcal{X}(i) \cup \mathcal{X}(i+1))$ that depends on the value of $X^i_{(j)}$ and on the values of the other variables in $\mathcal{X}(i)$ and in $\mathcal{X}(i+1)$.[24] For any $x \in (a_i, a_{i+1})$, let us write $\Delta^i(x)$ as short-hand for $\Delta^i(x, \mathcal{X}(i) \cup \mathcal{X}(i+1))$.

Now for each $i < A(\delta) - 1$ let $M^i_n(x) := \sum_{j=1}^{|\mathcal{X}(i)|} \mathbf{1}(nS^i_j \leq x)$ and for $i = A(\delta) - 1$, $M^i_n(x) := \sum_{j=1}^{|\mathcal{X}(i)|-2} \mathbf{1}(nS^i_j \leq x)$. For any interval i that is not the right-most subinterval the value $M^i_n(x)$ is the number of spacings scaled by n that are in (a_i, a_{i+1}) or that straddle (a_i, a_{i+1}) and (a_{i+1}, a_{i+2}) and whose length is less than x. For the right-most interval the value $M^i_n(x)$ obviously does not include spacings straddling an adjacent subinterval to the right. Now for each j such that $X_j \in \mathcal{X}(i)$ define the random variable

$$K_j(x, n, \Delta^i) := \begin{cases} 1 & \text{if at most one random variable in } \mathcal{X} \text{ falls in } \left(X_j, X_j + \frac{x}{n} \frac{2}{C'(X_j) + \Delta^i(X_j)}\right] \\ 0 & \text{otherwise} \end{cases}.$$

---

[24] For the right-most subinterval ($i = A(\delta) - 1$) the function Δ^i depends only on the values of the random variables in that subinterval.

Following the argument given in the proof of Propostion 2 to obtain (9), for each $i \leq A(\delta) - 1$ we can write

$$M_n^i(x) = \sum_{j:X_j \in \mathcal{X}(i)} (1 - K_j(x, n, \Delta^i)) = |\mathcal{X}(i)| - \sum_{j:X_j \in \mathcal{X}(i)} K_j(x, n, \Delta^i).$$

If we now sum over all subintervals and divide by $n - 2$ we obtain

$$G_n(x) = \frac{1}{n-2} \sum_{i=0}^{A(\delta)-1} M_n^i(x) = \frac{n}{n-2} - \frac{1}{n-2} \sum_{i=0}^{A(\delta)-1} \sum_{j:X_j \in \mathcal{X}(i)} K_j(x, n, \Delta^i).$$

Now for any constant d define

$$\mathcal{K}_n(x, d) := \frac{n}{n-2} - \frac{1}{n-2} \sum_{i=0}^{A(\delta)-1} \sum_{j:X_j \in \mathcal{X}(i)} K_j(x, n, d) = \frac{n}{n-2} - \frac{1}{n-2} \sum_{j:X_j \in \mathcal{X}} K_j(x, n, d).$$

The second equality follows from the fact that for any constant d, $K_j(x, n, d)$ does not depend on any particular subinterval. Finally, by the construction of the intervals we know that $|\Delta^i(x)| < \delta$ for each x and each i, so we can now write

(17)
$$\mathcal{K}_n(x, -\delta) \leq G_n(x) \leq \mathcal{K}_n(x, +\delta).$$

To complete the proof it now suffices to show for $d \in \{-\delta, +\delta\}$ and for each $x \geq 0$ that $\lim_{d \to 0} \lim_{n \to \infty} \mathbb{E}[\mathcal{K}_n(x, d)] = G(x)$ and $\lim_{d \to 0} \lim_{n \to \infty} \text{Var}[\mathcal{K}_n(x, d)] = 0$. In other words, the upper and lower bounds in equation (17) each converge in probability to $G(x)$ when the sample size becomes large and when the mesh size over the subintervals becomes small, and by virtue of being squeezed between the bounds, the convergence of the value of interest $G_n(x)$ follows.

The rest of the proof follows steps that are similar to the ones given in the proof of Proposition 3. Notice that $\mathbb{E}[K_j(x, n, d)]$ is equal to

$$\text{Pr}\left(\text{at most one random variable in } \mathcal{X} \text{ falls in } \left(X_j, X_j + \frac{x}{n} \frac{2}{C'(X_j) + d} \right] \right)$$

(18)
$$= \int_{-\infty}^{\infty} (n-1) \{R_n(x, z, d)\} [1 - R_n(x, z, d)]^{n-2} + [1 - R_n(x, z, d)]^{n-1} \, dL_n(z),$$

where $R_n(x, z, d) := L_n(z + 2xn^{-1}(C'(z) + d)^{-1}) - L_n(z)$. Equation (18) is obtained in a manner that is similar to the way in which equation (15) was obtained. Recalling that the X_i are drawn independently from L_n for each i, equation (18) is derived as follows: conditional on $X_j = z$ the term in the first set of square brackets taken to the power of $n - 2$ is the probability that $n - 2$ of the variables X_k for $k \neq i$ fall outside the interval $(z, z + 2xn^{-1}(C'(z) + d)^{-1}]$ and the term in curly brackets is the probability that the remaining variable falls in the interval. The entire expression is multiplied by $n - 1$ to account for the number of ways that

this event can occur. The term in the second set of square brackets taken to the power of $n-1$ is the probability that none of the $n-1$ variables X_k for $k \neq i$ fall outside the interval.

Define $\gamma(x, n, d)$ as the expression given in (18). Using arguments similar to those given in the proof of Remark 5, which rely on $L_n \to L$, $L_n' \to L'$, and $L_n'' \to L''$ uniformly, one can show that

$$\lim_{n \to \infty} \gamma(x, n, d) = \int_{-\infty}^{\infty} \left(\frac{L'(z)2x}{C'(z) + d} + 1 \right) \exp \left\{ -\frac{L'(z)2x}{C'(z) + d} \right\} dL(z).$$

By Proposition 1, which is due to Osborne and Pitchik (1986), $L = C$. So we have that

$$\lim_{n \to \infty} \gamma(x, n, d) = \int_{-\infty}^{\infty} \left(\frac{C'(z)2x}{C'(z) + d} + 1 \right) \exp \left\{ -\frac{C'(z)2x}{C'(z) + d} \right\} dC(z).$$

It is then straightforward to see that

$$\lim_{d \to 0} \lim_{n \to \infty} \mathbb{E}[\mathcal{K}_n(x, d)] = 1 - \lim_{d \to 0} \lim_{n \to \infty} \gamma(x, n, d) = G(x).$$

By a similar reasoning, we find that for each d, $\lim_{n \to \infty} \mathbb{E}[(\mathcal{K}_n(x, d)^2] = \lim_{n \to \infty} \mathbb{E}[\mathcal{K}_n(x, d)]^2$, so $\lim_{d \to 0} \lim_{n \to \infty} \text{Var}[\mathcal{K}_n(x, d)] = 0$, which completes the proof. $\qquad \square$

Proof of Remark 4. I simply show that all firms drawing their locations from a uniform distribution on $[0, 1]$ is a symmetric mixed strategy Nash equilibrium when $n \to \infty$. The rest of the result follows directly from Remark 3.

The expected payoff for firm i when it locates at x while all other firms choose their location at random according to L_n is given by $\tilde{\pi}_i(x) := \sum_{i=1}^{5} W_i(x)$ where

$$W_1(x) := \binom{n-1}{2} \binom{2}{1} \int_x^1 \int_z^1 (1 - L_n(y))^{n-3} C \left(\frac{x+y}{2} \right) dL_n(y) dL_n(z),$$

$$W_2(x) := \binom{n-1}{1} \binom{n-2}{2} \binom{2}{1} \times$$
$$\int_0^x \int_x^1 \int_z^1 (1 - L_n(y))^{n-4} C \left(\frac{x+y}{2} \right) dL_n(y) dL_n(z) dL_n(z'),$$

$$W_3(x) := \sum_{k=2}^{n-3} \binom{n-1}{k} \binom{k}{2} \binom{2}{1} \binom{n-k-1}{2} \binom{2}{1} \times$$
$$\int_x^1 \int_0^x \int_z^1 \int_0^{z'} L_n(y')^{k-2} (1 - L_n(y))^{n-k-3} \times$$
$$\left[C \left(\frac{x+y}{2} \right) - C \left(\frac{x+y'}{2} \right) \right] dL_n(y') dL_n(y) dL_n(z') dL_n(z),$$
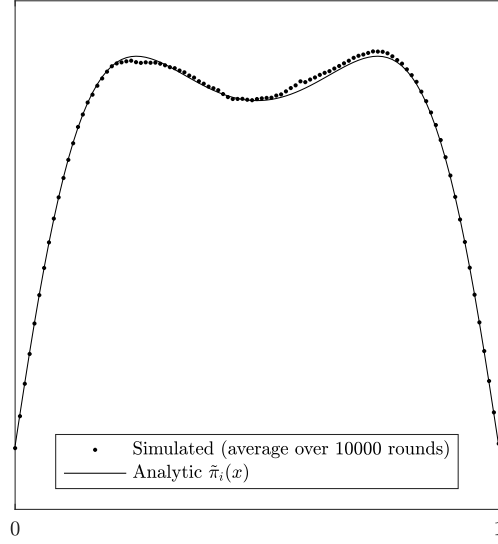
27

FIGURE 5. Illustration of $\tilde{\pi}_i(x)$. The number of firms is set to $n = 10$ and each point is the payoff at the relevant location averaged over 10000 simulated rounds.

$$W_4(x) := \binom{n-1}{1}\binom{n-2}{2}\binom{2}{1} \times$$

$$\int_x^1 \int_0^x \int_0^z L_n(y)^{n-4}\left[1 - C\left(\frac{x+y}{2}\right)\right] dL_n(y)dL_n(z)dL_n(z'),$$

$$W_5(x) := \binom{n-1}{2}\binom{2}{1}\int_0^x \int_0^z L_n(y)^{n-3}\left[1 - C\left(\frac{x+y}{2}\right)\right] dL_n(y)dL_n(z).$$

The equation for $\tilde{\pi}_i(x)$ is derived by considering all possible cases: $W_1(x)$ captures the case in which i locates at x such that $0 < x < z < y < (n-3$ variables$) < 1$; $W_2(x)$ captures the case in which i locates at x such that $0 < z' < x < z < y < (n-4$ variables$) < 1$; $W_3(x)$ captures the case in which i locates at x such that $0 < (k-2$ variables$) < y' < z' < x < z < y < (n-k-3$ variables$) < 1$, for k from 2 to $n-3$; $W_4(x)$ captures the case in which i locates at x such that $0 < (n-4$ variables$) < y < z < x < z' < 1$; and finally $W_5(x)$ captures the case in which i locates at x such that $0 < (n-3$ variables$) < y < z < x < 1$.

After employing the binomial theorem on $W_3(x)$, integrating by parts (multiple times), and then substituting $L_n(x) = C(x) = x$ for all x (so that the firms choose their locations

28

uniformly and the consumers are distributed uniformly), one obtains

$$\tilde{\pi}_i(x) = \frac{2}{n} - \frac{x^n}{n} - \frac{(1-x)^n}{n} + \frac{n-1}{2}\left[(1-x)^{n-2}x^2 + x^{n-2}(1-x)^2\right].$$

An illustration of this function is shown in Figure 5 when $n = 10$. It is easy to verify that $\lim_{n\to\infty} n\tilde{\pi}_i(x) = 2$ for all $x \in (0,1)$.[25] In other words, as the number of firms becomes large, the scaled payoffs become the same at each location. So when the distribution of consumers is uniform, the firms choosing their locations uniformly at random is a symmetric mixed equilibrium of the limiting game. □

APPENDIX B. FURTHER SIMULATIONS

Figure 6 repeats the simulations of subsection 3.1 for the uniform case for different values of n. The firm locations X_i are drawn from a uniform distribution on $[0,1]$ and the distribution of consumers is uniform. The histogram of the scaled market shares is shown in grey and the density of a Gamma$(2,2)$ distribution is shown as a solid black line. The panels show the outcomes for n equal to 50, 100, 250, and 1000. Clearly, the fit improves as n becomes larger. Since the strategy employed by the firms is the *limiting* one for each n, the speed with which the empirical distribution converges to the limiting distribution in these simulations may not accurately represent how fast the distribution would converge if the firms were playing an *equilibrium* strategy for each n. Following the discussion in subsection 3.1 I repeat, though, that only the limit matters for the purpose of illustrating Theorem 1.

Figure 7(a) illustrates Remark 1. I chose $F(x) = x$ and $V(x) = x^2$ for all $x \in [0,1]$. In other words, the gold panners locate along the river uniformly at random and since $v(x) = 2x$ there is twice as much gold at the end of the river than at the source. The histogram of the scaled values of gold collected across panners is shown in grey (with $n = 1000$) and the density derived from equation (5) is shown as a solid black line.

Figure 7(b) illustrates Remark 4. I set the distribution of consumers to be uniform on $[0,1]$ and let $n = 1000$ firms each draw their location independently from this distribution. The histogram of the scaled payoffs is shown in grey (with $n = 1000$) and the density of a Gamma$(4,2)$ distribution is shown as a solid black line.

---

[25]Performing similar steps on $\pi_i(x)$ in equation (1) (namely, evaluating the equation when $L_n(x) = C(x) = x$ for all x), one obtains that $\pi_i(x) = \frac{1}{n} - \frac{1}{2}\frac{x^n}{n} - \frac{1}{2}\frac{(1-x)^n}{n} + \frac{1}{2}\left[(1-x)^{n-1}x + x^{n-1}(1-x)\right]$. It is then easy to see that $\lim_{n\to\infty} n\pi_i(x) = 1$ for all $x \in (0,1)$.
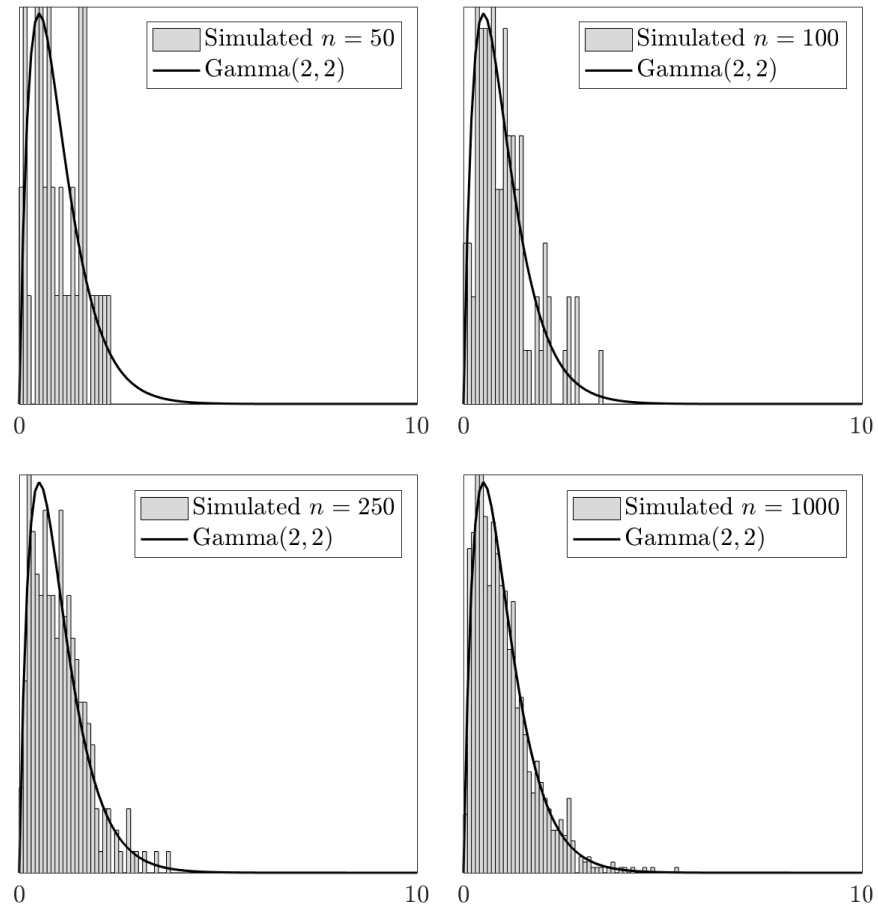
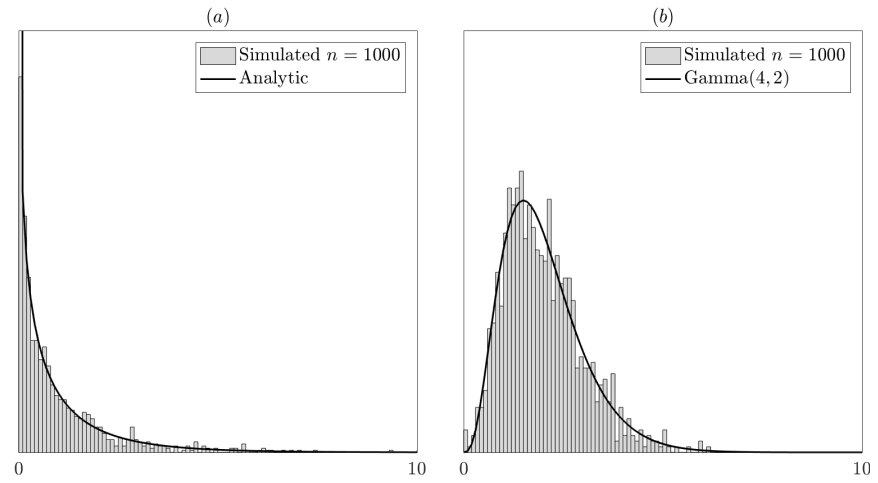FIGURE 6. Simulation results in the uniform case, varying n.

FIGURE 7. Illustration of Remark 1 and Remark 4.

References

Angle, J. (1993). Deriving the size distribution of personal wealth from "the rich get richer, the poor get poorer". *The Journal of Mathematical Sociology 18*(1), 27–46.

Ansari, A., N. Economides, and A. Ghosh (1994). Competitive positioning in markets with nonuniform preferences. *Marketing Science 13*(3), 248–273.

Boppana, M., R. Hod, M. Mitzenmacher, and T. Morgan (2016). Voronoi choice games. *arXiv preprint arXiv:1604.07084*.

Chen, Y., J. K. Lai, D. C. Parkes, and A. D. Procaccia (2013). Truth, justice, and cake cutting. *Games and Economic Behavior 77*(1), 284–297.

Cooper, R. B. (1981). *Introduction to Queueing Theory*. North Holland,.

Dasgupta, P. and E. Maskin (1986a). The existence of equilibrium in discontinuous economic games, i: Theory. *The Review of Economic Studies 53*(1), 1–26.

Dasgupta, P. and E. Maskin (1986b). The existence of equilibrium in discontinuous economic games, ii: Applications. *The Review of Economic Studies 53*(1), 27–41.

David, H. A. and H. N. Nagaraja (2003). *Order Statistics*. Wiley, New York.

De Palma, A., V. Ginsburgh, Y. Y. Papageorgiou, and J.-F. Thisse (1985). The principle of minimum differentiation holds under sufficient heterogeneity. *Econometrica*, 767–781.

Eaton, B. C. and R. G. Lipsey (1975). The principle of minimum differentiation reconsidered: Some new developments in the theory of spatial competition. *The Review of Economic Studies 42*(1), 27–49.

Ewerhart, C. (2015). Mixed eequilibrium in a pure location game: The case of $n \geq 4$ firms. *The BE Journal of Theoretical Economics 15*(2), 457–472.

Fournier, G. (2016). General distribution of consumers in pure hotelling games. *arXiv preprint arXiv:1602.04851*.

Gabaix, X. (1999). Zipf's law for cities: an explanation. *The Quarterly Journal of Economics 114*(3), 739–767.

Gabaix, X. (2009). Power laws in economics and finance. *Annual Review of Economics 1*(1), 255–294.

Gabaix, X. and A. Landier (2008). Why has ceo pay increased so much? *The Quarterly Journal of Economics 123*(1), 49–100.

Gabszewicz, J. J. and J.-F. Thisse (1992). Location. *Handbook of Game Theory with Economic Applications 1*, 281–304.

Gibrat, R. (1931). *Les Inégalités Économiques*. Recueil Sirey.

Holst, L. (1980). On the lengths of the pieces of a stick broken at random. *Journal of Applied Probability 17*(3), 623–634.

Hotelling, H. (1929). Stability in competition. *The Economic Journal 39*(153), 41–57.

Kleiber, C. and S. Kotz (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*, Volume 470. John Wiley & Sons.

Laster, D., P. Bennett, and I. S. Geoum (1999). Rational bias in macroeconomic forecasts. *The Quarterly Journal of Economics 114*(1), 293–318.

Loertscher, S. and G. Muehlheusser (2011). Sequential location games. *The RAND Journal of Economics 42*(4), 639–663.

Malueg, D. A. (2010). Mixed-strategy equilibria in the nash demand game. *Economic Theory 44*(2), 243–270.

Neven, D. J. (1986). On hotelling's competition with non-uniform customer distributions. *Economics Letters 21*(2), 121–126.

Núñez, M. and M. Scarsini (2017). Large spatial competition. In *Spatial Interaction Models*, pp. 225–246. Springer.

Osborne, M. J. and C. Pitchik (1986). The nature of equilibrium in a location model. *International Economic Review*, 223–237.

Ottaviani, M. and P. N. Sørensen (2006). The strategy of professional forecasting. *Journal of Financial Economics 81*(2), 441–466.

Prescott, E. C. and M. Visscher (1977). Sequential location among firms with foresight. *The Bell Journal of Economics*, 378–393.

Pyke, R. (1965). Spacings. *Journal of the Royal Statistical Society. Series B (Methodological)*, 395–449.

Ronayne, D. (2016). Extreme idealism and equilibrium in the hotelling-downs model of political competition. Technical report, Centre for Research in Economic Theory and its Applications CRETA.

Rudin, W. (1976). *Principles of Mathematical Analysis* (3 ed.). McGraw-hill New York.

Shaked, A. (1982). Existence and computation of mixed strategy nash equilibrium for 3-firms location problem. *The Journal of Industrial Economics*, 93–96.

Sutton, J. (1998). *Technology and Market Structure*. MIT Press.

Weiss, L. (1955). The stochastic convergence of a function of sample successive differences. *The Annals of Mathematical Statistics*, 532–536.

Williams, H. C. and M. Senior (1977). A retail location model with overlapping market areas: Hotelling's problem revisited. *Urban Studies 14*(2), 203–205.

Xefteris, D. (2016). Stability in electoral competition: A case for multiple votes. *Journal of Economic Theory 161*, 76–102.

University of Oxford
Email address: `bassel.tarbush@economics.ox.ac.uk`