

Spatial Choice Processes and the Gamma Distribution

Hannes Malmberg

Kandidatuppsats 2011:3
Matematisk statistik
Juni 2011

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm



Mathematical Statistics
Stockholm University
Bachelor Thesis **2011:3**
<http://www.math.su.se>

Spatial Choice Processes and the Gamma Distribution

Hannes Malmberg*

Juni 2011

Abstract

From commuting literature we know that the distribution of distance to spatial choices often is unimodally distributed with skew to the right; often the distribution is close to a Gamma distribution. The pattern appears robust. This paper constructs a mathematical statistical choice model to account for this fact. The key idea is that there is a tension between costs increasing with distance parallel with the number of offers per radial segment increasing with distance. The first effect induces short-distance choices whereas the second gives incentives to choose options located far away. We show by simulation that these two effects are likely to give rise to Gamma distributed choice patterns under a wide array of distributional assumptions. After this, a more general theory for stochastic optimization in space is outlined where offers of random value are distributed randomly in space. Approximations are made using extreme value theory, and under these approximations, the result is proved analytically for exponentially distributed offers. The ergodicity of extreme value distributions suggests why observed patterns seem robust.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden.
E-mail: per.hannes.malmberg@googlemail.com . Supervisor: Ola Hössjer.

Contents

1	Introduction	1
2	Relation to Previous Theory	1
3	Empirical Patterns	2
4	Sketch Theory and Simulation	3
5	General Theory	5
5.1	The Max Measure	6
5.2	Random Offers and their Spatial Distribution	8
6	Approximations of Theory	9
6.1	A Primer on Extreme Value Theory	10
6.2	Small Interval Approximation	10
6.3	Exponential Distribution – Simplifying the Approximation . .	12
6.4	Exponential Offer Distribution and Gamma Choice Variable .	13
7	Simulation Validation of Theoretical Results	15
8	Conclusion and Further Developments	16

Acknowledgements

First I would like to thank my supervisors Ola Hössjer from the Department of Mathematics and Yves Zenou from the Department of Economics for their support when writing this essay. Insightful input from different disciplines have been very helpful when developing my ideas. I am also very grateful for the cooperativeness of the Departments of Mathematics and the Department of Economics in making this interdisciplinary thesis possible.

I also want to thank John Östh at the Department of Economic and Social Geography for providing data and for being a great conversationalist when it comes to research topics. Not the least I also want to thank Bo Malmberg who gave me the idea of using extreme value processes to model spatial choice many years ago, and who always has been ready to discuss possible and impossible ideas throughout my life.

1 Introduction

This paper concerns spatial choice processes and proposes a model that explains stylized facts of travel distance distributions.

Spatial choice processes abound in everyday life: we have to choose how far away from our work we should live; we have to decide how far to go for a restaurant or how far away to make friends. Spatial choice processes also appear in nature, where animals have to make decisions on how far to forage. There are countless examples. A proper understanding of how these processes work is important for many applications. In social network theory, all agents are embedded in space and how their linkage behavior depends on distance is crucial for the shape of the network. In environmental and planning policy decisions it is important to know how travel decisions are affected by changes in oil prices and infrastructure.

These processes also provide an interesting research topic given that there are clear stylized facts to be explored. In particular, it has been noted that the distribution of choice distances often is unimodally distributed with a skew to the right; a Gamma or lognormal distribution has been proposed as a description of the pattern. In this paper I also perform an empirical study of school commutes in Sweden which reveals a similar pattern.

Stable empirical relations over time, space and different applications call for a clear theory. A Gamma distribution has been derived through search theory or discrete choice theory by economists, but this paper aims for reducing the number of restricting assumptions to a minimum to explain observed patterns. The basic idea of my theory is that the distribution emerges through the tension between two different forces: that cost and area per radial segment both increase with distance. Simulations show that this is sufficient to give a Gamma distribution under linear costs for a wide array of distributional assumptions. We develop a theory to study this type of process analytically and prove convergence under some approximations in the case of exponentially distributed offers. The proof suggests that the method might be extended to other cases.

2 Relation to Previous Theory

The standard theory in economics for explaining spatial choices is discrete choice theory. The theory posits an additive disturbance term in the valuation of all alternatives and then derives the probability distribution over

different choices. The workhorse model is the conditional multinomial logit (see for example McFadden and Manski [2]), which says that an alternative with properties x has value

$$U_i(x) = V(x; \beta) + \epsilon_i$$

where β is a parameter vector to be estimated, and the error terms ϵ_i are independent and identically distributed according to a Gumbel distribution. Under this model the distribution of final choices given that choices have characteristics $\{x_i\}$ is given by:

$$P(i|x_i) = \frac{e^{V(x_i; \beta)}}{\sum_j e^{V(x_j; \beta)}}$$

Our final model will be equivalent to the multinomial model when the intrinsic valuation of choosing an option at distance r away is

$$V(r; \beta) = -cr + \log r$$

Nevertheless, our model will be us a different methodologically. We will not postulate a Gumbel distribution of our disturbance term. Instead, we will postulate a random distribution of offers in space with a general disturbance term. When my disturbance term is exponential, the multinomial logit model will be the limiting distribution, but our modeling framework allows for a greater heterogeneity in parameters and models, which makes the model easier to fit to data.

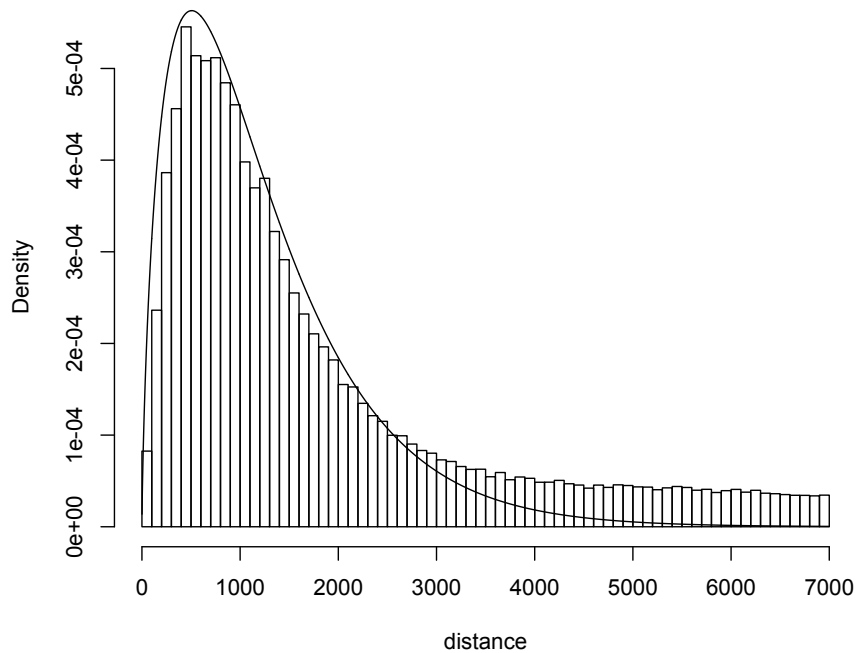
3 Empirical Patterns

It is an observed empirical fact that commuting distributions, a typical example of spatial link formation, often are unimodally distributed with skew to the right. A Gamma distribution or lognormal distribution are posited as good explanations of this pattern. Given this observation there have also been attempts in the economic literature to explain this. Rouwendal and Rietveld [4] showed that a Gamma distribution could be the result if we used a search theoretic approach and assumed that wage offers are exponentially distributed. Van Ommeren [3] showed that it is possible to get a unimodal distribution if we assume two-dimensional space and heterogeneity in jobs..

To get different examples than the labor market we have also looked how far Swedish school children commute. We use the unique database PLACE which contains the coordinates of all work, school and living positions in Sweden. From this, we look at 255,169 Swedish students who have a commute shorter than 7 km to school.

The data clearly exhibits the characteristic structure of a unimodal distribution with skew to the right. Another interesting question is whether school commutes follows a Gamma distribution as has been observed for job commuting. Figure 1 shows a Gamma curve fitted by hand over the distribution, which seems to produce a good fit for students traveling up to 3000 meters: a population of above 200,000 children, whereas the tail is too heavy afterwards.

Figure 1: Distribution of short school travel distances



4 Sketch Theory and Simulation

There is a relatively stable empirical pattern, and our hypothesis is that this should arise from two counteracting tendencies. First, costs are increasing with distance, which provides an incentive to choose options close to us. On the other hand, if we consider an interval $[r, r + \delta]$, we note that this interval has an approximate population of $n \approx 2\pi\rho r\delta$, where ρ is population density. Therefore, the number of offers per distance element, and therefore

the probability of receiving a good offer, increases with distance. Given that people do not commute an unbounded distance, the cost tendency should overcome the increasing population per radial element tendency in the limit. However, very few people commute extremely short distances, so the effect of increasing population per radial segment also seems to affect the qualitative structure of the outcome.

We explore the consequences of this simple setup by simulation. We make three assumptions:

- Offers have a random value and the number of offers per radial segment increases linearly in radius
- There is a linear cost cr of overcoming distance r
- We select the alternative which, net of transport costs, offers us the highest payoff

We simulate the choice distance as follows:

1. Generate i.i.d random variables $\{X_i^{(r)} : r \in \{1, 2, \dots, 1000\}, 1 \leq i \leq r\}$ from distribution F
2. Define $M_r = -cr + \max_{1 \leq i \leq r} \{X_i^{(r)}\}$ where c is a cost parameter.
3. Return j such that $M_j \geq M_k$ for all $k \in \{1, 2, \dots, 1000\}$.

Some notes. Firstly, we decide to generate r random variables for distance r . This corresponds to the special case $\rho = 1/(2\pi)$ but can be generalized; the key is that our strategy captures that the number of offers per distance element increases linearly with radius. Secondly, as long as F is an absolutely continuous probability distribution, the probability of a tie is zero and therefore the third step almost surely will return a unique element.

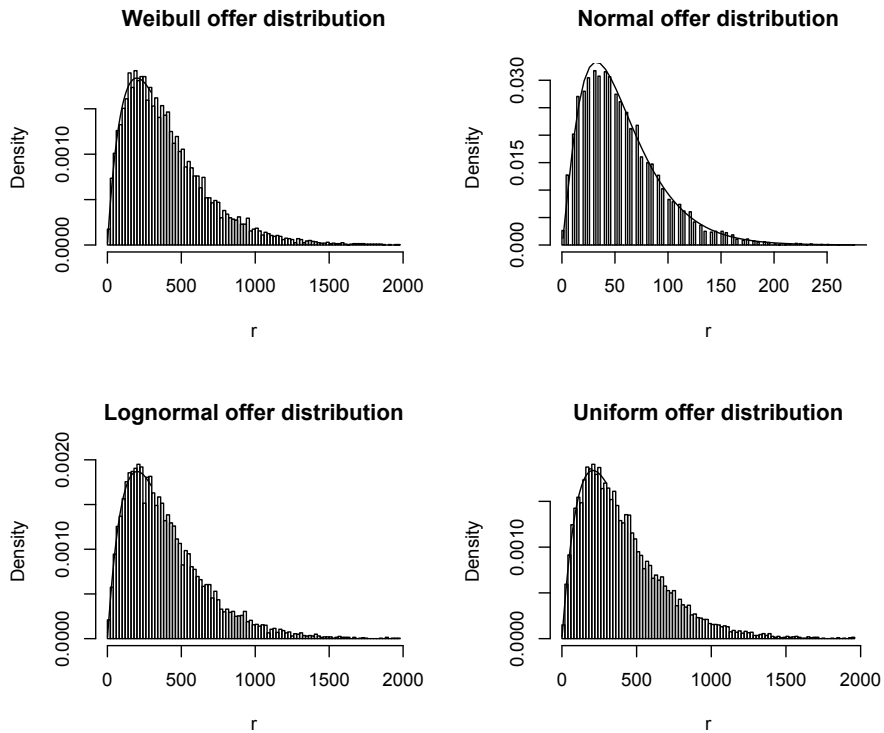
This simulation is performed for $c = 0.01$ and for the following choices of F : Normal, chi-square, lognormal, Weibull and uniform. Histograms, inspection and simple moment fitting showed that a Gamma-distribution is a reasonable approximation for all distributions of X . The results are shown in the figure. A Kolmogorov-Smirnov test failed to reject the null hypothesis that D_F was Gamma distributed with the parameters implied by a simple fitting of moments for all distributions except for the normal distribution.

1

¹An additional study of the normal distribution shows that the rejection of a gamma distribution can be an artifact of the construction of the simulation. We include a theoretical and empirical distribution curve for the normal case in Figure 3 and it appears that

To sum up, we conclude that under the model of utility maximization under linear cost, a wide array of offer distribution functions give rise to a choice distance distribution which is Gamma distributed.

Figure 2: Simulations and theoretical choice distance distributions for a number of different distributions F

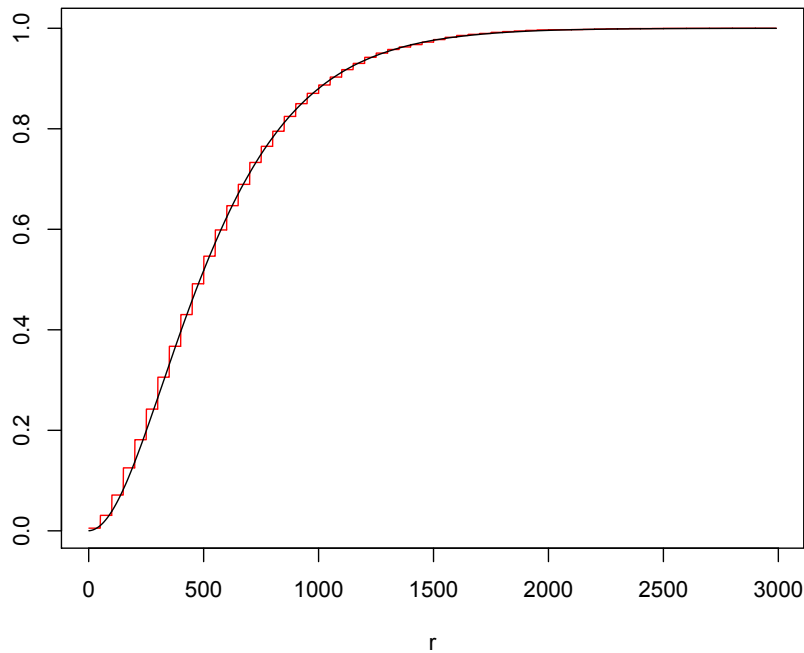


5 General Theory

The robustness of the results in the simulations seems to suggest that there should be a theoretical reason for why this occurs. However, in order to prove analytical results, we have to state the problem more precisely. Our problem is to construct a continuous random variable D that captures at what distance our preferred choice will be located. We build a model where

the spacing of r by 1 which by design gives rise to the error for small values, and means that the Kolmogorov-Smirnov test statistic is larger than could be expected by chance. Therefore, we cannot guarantee that the rejection is not due to the particular design of our simulation. The exact nature of how the normal offer distribution and the Gamma distribution relate is a point of further research.

Figure 3: Theoretical (black) and empirical (red) distribution functions for normally distributed F



we first define in general terms what is meant by a continuous random choice process. The key concept in this section will be the "max-measure" which defines how we make choices over a finite number of disjoint intervals when valuations are random. We show how a continuous random choice variable can be seen as an induced measure of this max measure.

In the second section we give the problem more structure by assuming that the (random) value of an interval can be seen as a maximization of primitives called "offers". We give a formal definition of what an offer is and what our model is for their spatial distribution.

5.1 The Max Measure

Definition 1. Let $\mathcal{I} = \{[a, b) : [a, b) \subset [0, T]\}$ and let $X : \mathcal{I} \rightarrow L^1$, where L^1 is the space of integrable random variables. We call X a max measure if the following five properties hold:

- The max measures of disjoint intervals are independent
- If $I = A \cup B$ we have $X(I) = \max\{X(A), X(B)\}$
- $X(I)$ is absolutely continuous for all intervals I
- If $A_1 \supset A_2 \dots$, and $\mu(A_n) \rightarrow 0$, where μ is the Lebesgue measure, then $X(A_n) \rightarrow -\infty$ in probability.
- $X(\emptyset) = -\infty$

Remark 2. The second requirement is meaningful as it defines the limit of a non-decreasing sequence of random variables. The third requirement is a technical requirement to ensure that ties only occur with probability 0.

Lemma 3. Let \mathcal{A}_T be the class of finite unions of disjoint subintervals of $[0, T]$ and suppose we have a max measure X with the property that $P(X([0, T]) > -\infty) = 1$. Consider the set function F defined by on \mathcal{A}_T

$$F(A) = P(X(A) \geq X([0, T]))$$

This set function extends to a unique measure on the positive real line, which is called an arg-max measure.

Proof. It is immediate that \mathcal{A}_T is a ring on $[0, T]$. Furthermore we have that $F(A) \in [0, 1]$ and $F([0, T]) = 1$. If we can prove countable additivity we can use Carathéodory's extension theorem to conclude that we have a unique measure on $[0, T]$.

The first step is to establish that F is finitely additive on \mathcal{A}_T . I.e. if $A = \cup_{i=1}^n A_i$, where the A_i 's are disjoint, we have

$$F(A) = \sum_{i=1}^n F(A_i)$$

The easiest way to show this is by equating events in the probability space.

$$\{X(\cup_{k=1}^n A_k) \geq X([0, T])\} = \cup_{k=1}^n \{X(A_k) \geq X([0, T])\}$$

where the elements on the right-hand side are disjoint. Indeed, the definition of the max measure gives us $X(\cup_{k=1}^n A_k) = \max_k X(A_k)$. Therefore we know that if the max measure of a union is larger than or equal to the maximum over the whole interval, it is equivalent to the max measure of one its components being larger than or equal to the maximum over the whole interval.

We deduce disjointness by noting that

$$X([0, T]) = \max\{X(A_i), X([0, T] \setminus A_i)\} \geq X(A_i)$$

From the observation above, we note that

$$\{X(A_i) \geq X([0, T])\} \Rightarrow X(A_i) = X([0, T])$$

Hence, if we are on the intersection of $\{X(A_i) \geq X([0, T])\}$ and $\{X(A_j) \geq X([0, T])\}$ we have $X(A_i) = X(A_j)$. But $X(A_i)$ and $X(A_j)$ are both independent and absolutely continuous and therefore their difference is also absolutely continuous. Therefore, their difference will be zero with probability zero. Hence, the sets $\{X(A_i) \geq X([0, T])\}$ and $\{X(A_j) \geq X([0, T])\}$ are disjoint with probability 1, and we can use additivity of probability measures to conclude that:

$$F(\cup_{i=1}^n A_i) = \mathbb{P}[X(\cup_{i=1}^n A_i) \geq X([0, T])] = \sum_{i=1}^n \mathbb{P}(X(A_i) \geq X([0, T])) = \sum_{i=1}^n F(A_i)$$

Therefore, we know that F is a finitely additive set function on \mathcal{A}_T . To prove countable additivity we have to show that if we have a decreasing chain of subsets

$$A_1 \supset A_2 \supset A_3 \dots$$

such that $\cap_n A_n = \emptyset$, then $F(A_n) \rightarrow 0$. The technical assumptions we have made on the max measure X make the result simple. Indeed, if $\cap_n A_n = \emptyset$, we have that $\mu(A_n) \rightarrow 0$ there μ is the Lebesgue measure. This means that

$$X(A_n) \rightarrow -\infty$$

in probability. But by definition $X([0, T]) > -\infty$ with probability 1, and hence

$$F(A_n) = P(X(A_n) \geq X([0, T])) \rightarrow 0$$

as $n \rightarrow \infty$ which completes the proof. \square

5.2 Random Offers and their Spatial Distribution

The previous section defined a max measure which assigns a random variable to each interval and a corresponding max measure measure. In this section we introduce a more basic random variable called *offer*, which can be used to derive a particular argmax measure. The offers are located in space as a homogenous Poisson process, and the value of the offer is a random component minus the cost of reaching it, which is linear in the distance required to reach the offer.

We begin by defining the distribution of offers on the real line. Consider

$$N(x) \sim PP(\pi \rho x^2)$$

where this means that $N(x)$ is an inhomogenous Poisson process on the positive real line with local intensity $2\pi\rho x$. This is simply the radial version of a homogenous Poisson process with intensity ρ on the plane. We denote the corresponding jump process $\{J_n\}$, and connect this Poisson process with offers by considering a sequence of random variable pairs

$$(J_n, O_n)$$

where $O_n = -cJ_n + \epsilon_n$, and where the ϵ_n are absolutely continuous, independent and identically distributed random variables with common probability distribution F .

These two definitions give us the tool to define our max measure. Let $A \subset [0, T]$ be a finite interval. Then we can define an arg-max measure by the following rule:

$$X(A) = \max\{O_n : J_n \in A\}$$

and $X(A) = -\infty$ if $\{J_n : J_n \in A\}$ is empty. This means that the random value of an interval A is simply the value of the best offer available in that interval, if there is an offer available. This is a max measure. Indeed, if A and B are disjoint, $X(A)$ and $X(B)$ will be independent as the Poisson process has independent increments and the random components in the offers are independent. Absolute continuity follows from the corresponding property of the ϵ_i 's and if the measure of A_n tends to 0, the probability that the interval will be empty, and therefore have value $-\infty$, tends to 1. It is clear that $X(A \cup B) = \max\{X(A), X(B)\}$.²

6 Approximations of Theory

The model outlined in the section above gives us a theoretical base for exploring questions about the continuous distribution of choices. The Poisson structure means that one strategy could be to solve for the choice random variable directly by simply conditioning on the number of offers and derive the probability distribution of the maximum offer directly by using the fact that the jumps of a Poisson process are uniformly distributed conditioned on the number of jumps.

²To apply Lemma 4 we also need that $X([0, T]) > -\infty$ *a.s.*. In the original formulation this is not true as there might be 0 jumps in $[0, T]$. To be very formal, this can most likely be solved by defining a limit $X(\mathbb{R}^+) \equiv \lim_{t \rightarrow \infty} X([0, t])$ which exists almost surely. On this we have one jump almost surely and we can translate the theory to this setting. Alternatively we can introduce an extra offer which is randomly distributed and make sure that the theory is not hurt by the degree of interdependence introduced by this extra offer.

However, there are some drawbacks to this strategy. Firstly, the tractability of the problem would depend to a large extent on the distribution F of the random terms in the offers. Secondly, the conditioning procedure and integrating over possible locations of the maximum offer is not straightforward. We will therefore use another integration strategy. Using a succession of simplifying assumptions, we will derive approximate probabilities of choosing an alternative from a small interval $[r, r + \delta]$. Thereafter, we can use the additivity of choice probabilities of disjoint sets to get an integral approximation of the probability of choosing an alternative in a particular interval.

6.1 A Primer on Extreme Value Theory

Our approximations will depend a lot on the property of extreme value distributions. Extreme value distribution theory describes the asymptotic behavior of maxima of collections of independent random variables. The benefit of extreme value theory is its ergodicity properties: in the limit, the distribution of the maximum has a very weak dependence on the initial distribution. This robustness makes them a useful tool in stochastic optimization applications. Extreme value theory has been used previously in discrete choice theory, but this paper takes a somewhat different approach from that literature: an extreme value distribution is not postulated, but derived as a consequence of more basic assumptions. The main result from extreme value theorem that we will use is the following:

Theorem 4 (Fisher-Tippet-Gnedenko Theorem (Extreme Value Theorem)). *Let $\{X_n\}$ be a sequence of independent and identically distributed random variables and let $M_n = \max\{X_1, X_2, \dots, X_n\}$. If there exist sequences $\{a_n\}$ and $\{b_n\}$ with $a_n > 0$ such that:*

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = H(x)$$

then $H(x)$ belongs to either the Gumbel, the Frechét, or the Weibull family. [1]

Remark 5. *Under a wide range of distributions of X_n , convergence does occur, and for most common distributions the convergence is to the Gumbel distribution, which has the form $Gumbel(x) = \exp\left(-\exp\left(-\frac{x-\mu}{\beta}\right)\right)$ for some parameters μ, β .*

6.2 Small Interval Approximation

We now turn to approximation. First, we use a mean value approximation of the number of offers in a particular interval. Neglecting quadratic terms

in δ , we can approximate the number of offers in an interval $[r, r + \delta]$ with $2\pi r\rho\delta$. The value of an interval is the best of these different offers, and we can approximate this applying extreme value theory to the disturbance term ϵ_i .

Suppose that $M_n = \max\{X_1, \dots, X_n\}$ where the X_i 's are identically and independently distributed according to F . We assume that F belongs to a class of distributions such that there exist sequences $\mu_n(F)$ and $\beta_n(F)$ where

$$\mathbb{P}\left(M_n \leq \frac{x - \mu_n(F)}{\beta_n(F)}\right) \rightarrow \text{Gumbel}(x; 0, 1)$$

where

$$\text{Gumbel}(x; \mu, \beta) = \exp\left(-\exp\left(-\frac{x - \mu}{\beta}\right)\right)$$

is the Gumbel distribution. We note that this means that for large n , we have

$$\mathbb{P}(M_n \leq x) \approx F(x; \mu_n(F), \beta_n(F))$$

We write $\mu_n(F)$ and $\beta_n(F)$ to denote that the form of convergence depends on the original probability distribution.

The problem of applying this theory to our case is that our mean value approximation often means that we have a non-integer number of random variables in a particular interval, $2\pi\rho r\delta$ namely. However, in many cases, $\mu_n(F)$ and $\beta_n(F)$ has a functional form such as $a + b \log(n)$ for large n . This means that we get a clear interpretation of taking the maximum over a non-integer number of random variables. If we neglect the variation in cost over small intervals, this leads us to the following definition of our mean-value, extreme value approximation.

Definition 6 (Mean-value/Extreme-value approximation). *Let J_n be the jump process of an inhomogenous Poisson process with local intensity $2\pi r\rho$ and let $O_n = -cJ_n + \epsilon_n$ be a sequence of random variables where the ϵ_i are absolutely continuous, independent random variables distributed according to a common probability distribution F . If we define $X(A) = \max\{O_n : J_n \in A\}$ and F belongs to the class of probability distribution whose extreme value statistic converges to the Gumbel distribution for some normalizing sequences $\mu_n(F)$ and $\beta_n(F)$ which have an asymptotic functional form $m_F(n)$ and $b_F(n)$ respectively, then we make the following approximation for small intervals:*

$$X([r, r + \delta]) = -cr + G_r$$

where $G_r \sim \text{Gumbel}(m_F(2\pi\rho r\delta), b_F(2\pi\rho r\delta))$

6.3 Exponential Distribution – Simplifying the Approximation

The approximation above gives us a strong tool to derive the arg-max measure for many initial distributions F . For simplicity, we will restrict our attention in this paper to the exponential distribution as the limiting parameter functions m_F and b_F take a particularly simple form under the exponential distribution. Under the exponential distribution, we will be able to show explicitly that the continuous choice process is a Gamma distribution.

Lemma 7. *Let $X_i \sim \text{Exp}(1)$, and $M_n = \max\{X_1, \dots, X_n\}$. Then $\mathbb{E}M_n = H(n)$, where $H(n)$ is the harmonic function.*

Proof. We note that if $X \sim \text{Exp}(1)$, $F(X) = 1 - e^{-X} \sim U[0, 1]$. The density of the maximum random variable M_n is $ne^{-x}(1 - e^{-x})^{n-1}$. Hence

$$\begin{aligned} \mathbb{E}M_n &= \int_0^\infty nx(1 - e^{-x})^{n-1}e^{-x} dx \\ &= n\mathbb{E}(X(1 - e^{-X})^{n-1}) \\ &= -n\mathbb{E}(\log(1 - U)U^{n-1}) \end{aligned}$$

where $U \sim [0, 1]$. We then note that:

$$\begin{aligned} \mathbb{E}M_n &= -n \int_0^1 \log(1 - x)x^{n-1} ds \\ &= -n \int_0^1 \left(-x - \frac{x^2}{2} - \frac{x^3}{3} \dots\right) x^{n-1} dx \\ &= n \int_0^1 \left(x^n + \frac{x^{n+1}}{2} + \frac{x^{n+2}}{3} + \dots\right) dx \\ &= \sum_{k=1}^{\infty} \frac{n}{k(n+k)} \\ &= \sum_{k=1}^{\infty} \frac{1}{k} - \frac{1}{n+k} \\ &= \sum_{k=1}^n \frac{1}{k} \\ &= H(n) \end{aligned}$$

□

Furthermore, for large n we have that:

$$\text{Var}M_n \approx \frac{\pi^2}{6}$$

For a general exponential variable with rate λ the corresponding quantities are $\frac{H(n)}{\lambda}$ and $\frac{\pi^2}{6\lambda^2}$. We know the first and second moments of the Gumbel distribution. If $G \sim \text{Gumbel}(\mu, \beta)$ we have

$$\begin{aligned}\mathbb{E}G &= \mu + \beta\gamma \\ \text{Var}(G) &= \frac{\beta^2\pi^2}{6}\end{aligned}$$

where γ is Euler's constant. Using that $\frac{H(n)-\log(n)}{\lambda} \rightarrow \frac{\gamma}{\lambda}$, we get $\beta_n \approx \frac{1}{\lambda}$ and $\mu_n \approx \frac{\log(n)}{\lambda}$. Therefore, using our mean-value/extreme-value approximation from the previous section, we define a small interval approximation of our arg-max measure as

$$X[r, r + \delta] \approx -cr + G_r$$

where $G_r \sim \text{Gumbel}(\frac{\log(2\pi\rho r\delta)}{\lambda}, \frac{1}{\lambda})$. The property of the Gumbel distribution means that we have that

$$\begin{aligned}y + \text{Gumbel}(\mu, \beta) &= \text{Gumbel}(y + \mu, \beta) \\ a \times \text{Gumbel}(0, \beta) &= \text{Gumbel}(0, a\beta)\end{aligned}$$

This means that

$$G_r = \frac{\log(2\pi\rho\delta)}{\lambda} + \frac{\log(r)}{\lambda} + \frac{1}{\lambda} \times G'_r$$

where $G'_r \sim \text{Gumbel}(0, 1)$. Identical, positive, affine shifts of collections of random variables do not change the max operation. Therefore, if we only compare intervals of the same length δ , we can simplify the problem to a standard form by the following affine transformation:

$$f : x \mapsto \lambda x - \log(2\pi\rho\delta) \Leftrightarrow f(X[r, r + \delta]) = -\lambda cr + \log(r) + G'_r = G_r^*$$

where $G_r^* \sim \text{Gumbel}(-c'r + \log(r), 1)$ where $c' = \lambda c$ is a rescaled cost parameter.

Once all the primes and stars are dropped, we have reduced the problem in case of an exponential distribution to that small intervals $[r, r + \delta]$ have valuation

$$G_r \sim \text{Gumbel}(-cr + \log(r), 1)$$

6.4 Exponential Offer Distribution and Gamma Choice Variable

The key result of this section is that if we have an exponential disturbance term on the offers, and therefore get the simplification above, the resulting

choice random variable (defined in Remark 5) is Gamma distributed. In the derivation we will assume that the Gumbel approximation holds even as the interval length becomes arbitrarily small. This is clearly unrealistic as the number of offers within each interval will decrease to the point where extreme value approximations are no longer feasible. Hence, the answer will be valid when the population density is sufficiently high so that we can apply extreme value theory to intervals small enough to make an integral approximation valid

Proposition 8. *Suppose that for all $r \geq 0$, $X[r, r + \delta] \sim \text{Gumbel}(-cr + \log(r), 1)$, $X(A \cup B) = \max\{X(A), X(B)\}$ and disjoint intervals are independent. Then for all intervals $A \subset [0, T]$ and all $T > 0$ we have $P[X(A) \geq X([0, T])] = \frac{P(G \in A)}{P(G \in [0, T])}$ where $G \sim \text{Gamma}(2, c)$*

Proof of Proposition. Let $R_\delta = \{t_0, t_1, \dots, t_{n+1}\}$ be a partition of $[0, T]$, with $0 = t_0 < t_1 < \dots < t_{n+1} = T$ and $t_{i+1} - t_i = \delta$ for all i . Let $r' \in R$ and write f_r and F_r for the density and distribution functions of $\text{Gumbel}(\log(r) - cr, 1)$. We will consistently write X_r to mean $X([r, r + \delta])$. Then:

$$\begin{aligned} \mathbb{P}(\max_{r \in R_\delta} X_r = X_{r'}) &= \int_0^\infty f_{r'}(x) \mathbb{P}(\max_{r \in R_\delta} X_r = X_{r'} | X_{r'} = x) dx \\ &= \int_0^\infty f_{r'}(x) \prod_{r \in R_\delta, r \neq r'} \mathbb{P}(X_r \leq x) dx \\ &= \int_0^\infty f_{r'}(x) \prod_{r \in R_\delta, r \neq r'} F_r(x) dx \end{aligned}$$

We now note that:

$$\begin{aligned} f_r(x) &= \frac{dF_r(x)}{dx} \\ &= \frac{d}{dx} \left(e^{-e^{-(x - \log(r) + cr)}} \right) \\ &= e^{-(x - \log(r) + cr)} F_r(x) \end{aligned}$$

Putting this result back in our previous expression gives us:

$$\begin{aligned} \mathbb{P}(\max_{r \in R_\delta} X_r = X_{r'}) &= \int_0^\infty f_{r'}(x) \prod_{r \in R_\delta, r \neq r'} F_r(x) dx \\ &= \int_0^\infty e^{-(x - \log(r') + cr')} \prod_{r \in R_\delta} F_r(x) dx \\ &= r' e^{-cr'} \int_0^\infty e^{-x} \prod_{r \in R_\delta} F_r(x) dx \end{aligned}$$

The important part of this result is that the expression within the integral sign is independent of r' . Therefore, the only dependence on r' is through $r'e^{-cr'}$. From the fact that we have a probability distribution summing to 1 we get:

$$\mathbb{P}(\max_{r \in R_\delta} X_r = X_{r'}) = \frac{r'e^{-cr'}}{\sum_{r \in R_\delta} r'e^{-cr}}$$

Now let us consider an interval $A \subset [0, T]$ and a sequence of partitions R_{δ_n} such that δ_n tends to 0. By assumption, the Gumbel-assumption should be true for all δ_n .

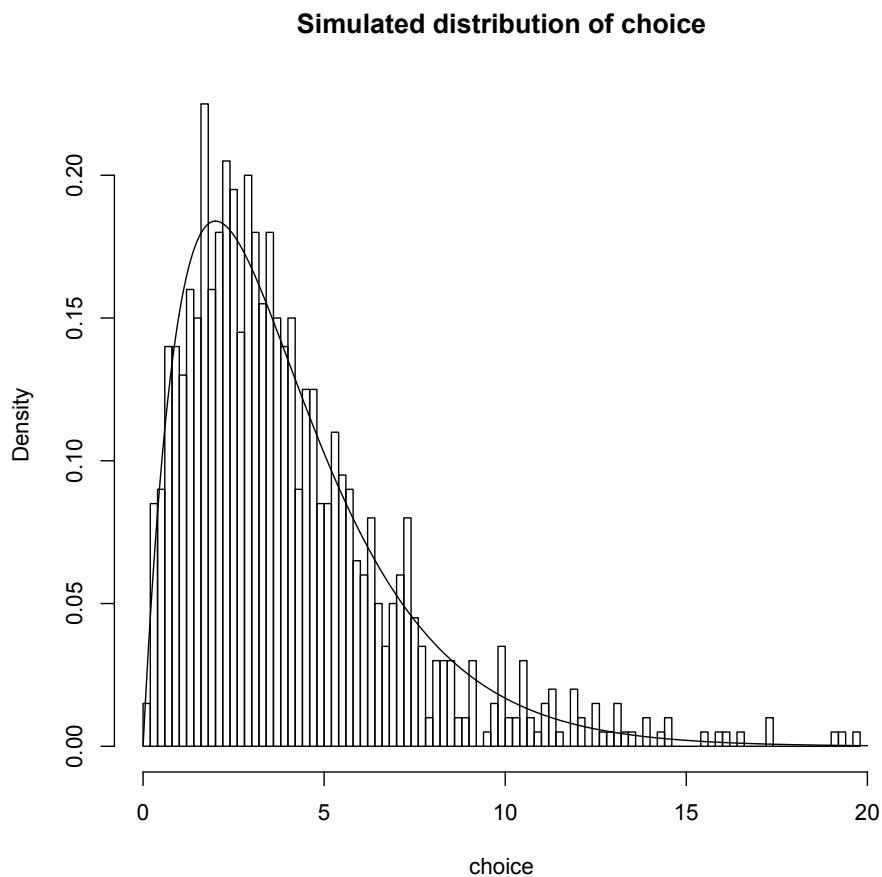
$$\begin{aligned} \mathbb{P}(\max_{r \in R_{\delta_n}} X_r \in A) &= \frac{\sum_{r' \in A \cap R_{\delta_n}} r'e^{-cr'}}{\sum_{r \in R_{\delta_n}} r'e^{-cr}} \\ &= \frac{\sum_{r' \in A \cap R_{\delta_n}} c^2 r' e^{-cr'} \delta_n}{\sum_{r \in R_{\delta_n}} c^2 r e^{-cr} \delta_n} \\ &\rightarrow \frac{\int_A c^2 r' e^{-cr'} dr'}{\int_0^T c^2 r e^{-cr} dr} \\ &= \frac{\mathbb{P}(G \in A)}{\mathbb{P}(G \in [0, T])} \end{aligned}$$

where the convergence comes from the definition of the Riemann integral. Hence, we obtain the required result \square

7 Simulation Validation of Theoretical Results

A lot of assumptions were made to derive the final result. The key assumption is that there is a sufficiently high population density so that we can apply extreme value theory, and it is interesting to see whether they are reasonable for the degree of population density observed in cities. We test this by simulation. We make a very crude approximation and note that median travel distance to work in Sweden is about 3km. If offers are distributed as $Exp(1)$ this corresponds to a travel cost parameter of $c \approx 0.5$ (using the theoretical result that the travel distance should be distributed as $Gamma(2, c)$). If we look at Stockholm a web crawler can find 22,000 job openings right now, so we consider a simple model where 50,000 offers with exponential disturbance terms are randomly distributed in a disk of radius 20km. We repeated the experiment 1,000 times and the result is presented in Figure 4 together with the predicted distribution. This shows that with a scaling leading to a median theoretical travel distance of 3km, we know that 50,000 offers are enough to ensure that our approximations work (Kolmogorov-Smirnov p-value: 0.9269).

Figure 4:



8 Conclusion and Further Developments

This paper shows that the simple idea of tension between area per radial segment and cost goes a long way of explaining observed stylized facts in spatial choice processes. The simulations showed that we could expect a Gamma distribution from a lot of offer distributions, and we managed to prove it theoretically for the exponential case. The paper also shows that a Gumbel distribution does not have to be postulated for discrete choice processes, but that it can arise naturally from limiting considerations.

There are a number of future developments that can be considered. Firstly, given the extremely close fit to a Gamma distribution for all simulations, it seems like we should be able to prove convergence for more distributions

than the exponential distribution. The problem with other distributions is that the dispersion parameter in the Gumbel distribution depends on the sample size in the limit. Simulation results suggest that this does not affect the results significantly, but it means that it is impossible to separate out r -dependence in the integral, which was a crucial step in the proof of Proposition 9. However, we can provide some simplification by observing that even if the dispersion parameter does depend on the sample size, the density function of a Gumbel distribution can be written as:

$$f_r(x) = a(r) \exp(-b(r)x) \times F_r(x)$$

where $F_r(x)$ is the distribution function. In our proof of Proposition 9 we considered the special case of $a(r) = r \exp(-cr)$ and $b(r) = 1$. Consider a partition R and let $X = \max_{r \in R} X_r$. Using a similar method to the one we used in the proof, and using integration by parts, we get that:

$$\begin{aligned} P(\max_{r \in R} X_r = X_s) &= a(s) \times \int_0^\infty e^{-b(s)x} F_X(x) dx \\ &= a(s)/b(s) * \int_0^\infty e^{-b(s)*x} f_X(x) dx \\ &= a(s)/b(s) * L_X(b(s)) \end{aligned}$$

where L_X is the Laplace transform of X . In the particular case of an exponential offer distribution, $b(s)$ is independent of s and the density is proportional to $r \exp(-cr)$. In the case of a more general distribution, we can use the fact that X is a maximum of Gumbel distributions to approximate it by a which can also be approximated by a Gumbel distribution. This method provides an indication for how one can provide numerical bounds for how the distribution will behave under different offer distributions. For example if changes in $b(s)$ are much smaller than changes in $a(s)$ we will approximately get a Gamma distribution.

Another improvement of the theory would be to introduce multidimensionality. Some empirical facts such as school choice are not likely to be the outcome of a single optimizing procedure. Instead, it is probably the result of a multidimensional optimizing procedure over job, residence, school and other amenities. If we look at the cross-section of school it has the unimodal distribution with skew to the right, and it would be good to see that in the joint distribution resulting from the multidimensional optimization procedure, the marginal distributions have the right properties.

In conclusion, this paper offers a highly general theory to account for stylized facts in spatial choice processes. Simulations give strong results, of which some can be proved analytically by verifiably reasonable assumptions. Lastly, there are numerous potential developments which means that the theory is unlikely to have reached its full potential.

References

- [1] GUMBEL, E.J., *Statistics of Extremes (new edition)*, Courier Dover Publications, 2004
- [2] MANSKI, C.F. & MCFADDEN, D., *Structural Analysis of Discrete Data with Econometric Applications*, Section III, MIT Press, 1981, <http://elsa.berkeley.edu/~mcfadden/discrete.html>
- [3] VAN OMMEREN, J, *The Commuting Distribution*, Discussion paper, Tinbergen Institute Discussion Paper, 2004
- [4] ROUWENDAL, J & RIETVELD, P, *Changes in Commuting Distances of Dutch Households*, *Urban Studies*, 1994 31:1545