Accessibility analysis for planning applications I: impedance functions

Anastasia Soukhov

The term **Accessibility** has many definitions. Within the context of transportation planning literature and practice, accessibility is often a location-based measure of *potential for interaction*. Specifically, it quantifies the potential of a population to reach opportunities in a given region based on their means of transportation. The “population” are the people/activities at an origin and the “opportunities” are the people/activities at a destination. For instance, the “population” might encompass individuals employed at a type of job if concerned with the potential access to jobs (the “opportunities), or children if measuring potential access to schools, or all people if quantifying potential access to all healthcare services. Different modes (e.g., walking, transit), time of travel, quality of route taken, and quality of”opportunities”, can all be considered.

The output of these measures are typically a value or normalized score that is assigned to each spatial unit (e.g., a census tract, neighbourhood boundary, parcel, etc.). This output provides a snapshot of the relationship between land-use and transportation in the region: areas with high scores are relatively well-connected and are in proximity to plenty of opportunities while the opposite is true for areas with low accessibility. Accessibility analysis can be used by planners to identify priority areas for transportation and “opportunity” improvements. For further reference, a unifying resource can be found in Levinson and King (2020) and a practical guide (using R) in Pereira and Herszenhut (2023).

Needless to say, accessibility analysis can be an important component when planning for equitable transportation and service provision.

This blog post is the first of a series of posts that present accessibility analysis for planning applications. The series will walk readers through the components of accessibility analysis as well as its potential uses when planning for equity. This post explores how travel behavior enters accessibility measures through the *impedance function*, the implications of travel behaviour assumptions and how one may select parameters for these assumptions. In the subsequent posts, I plan to discuss how these assumptions about travel behavior impact accessibility analysis outputs, different types of accessibility measures, and centering equity and justice conceptualizations in accessibility analysis.

## Counting opportunities based on travel behaviour assumptions

Many accessibility measures derive from (Hansen 1959), who was inspired by definitions of demographic potential interaction (Stewart 1948). This accessibility measure is represented in ([Equation 1](#eq-hansen-access)).

The accessibility score at each spatial unit is a weighted sum of the number of opportunities at . The weights in this summation are a function of the cost of travel, , sometimes called a distance-decay function. reflects how the potential for interaction changes with the cost of travel between spatial units and , that is the origin and destination of a potential trip. Generally, the more costly it is to reach a destination, the lower the potential for interaction.

The cost in the distance-decay function can be distance, time, financial cost, or a combination of several factors. Since distance is not always the unit of travel cost, is also known more generally as an *impedance function* since the function models the impedance of travel. Generally, declines with growing travel cost (the impedance is greater), and so opportunities at destinations that are less costly to reach are more heavily weighted in the summation that yields . Conversely, opportunities that are costly to reach (i.e., they are *far* away in terms of travel cost) have values of close to or equal to zero, so a negligible amount of enters the summation.

In short, the impedance function allows the accessibility analyst to precisely define a measure of travel behavior: the relationship between where people at an origin are and where they usually go, want to go or can go, to reach opportunities at destinations.

From this perspective, the definition of the impedance function is incredibly important. Let’s go over commonly defined impedance functions in accessibility research and their impact on opportunity-counting (the summation of opportunities) at specific travel costs , namely:

* Binary([Equation 2](#eq-binary-access))
* Uniform distribution([Equation 3](#eq-uniform-imped)})
* Exponential distribution([Equation 4](#eq-exp-imped))
* Gamma distribution([Equation 5](#eq-gamma-imped))

The **binary function** is shown in ([Equation 2](#eq-binary-access)). This function forms the basis of the cumulative opportunities measure approach (this measure is to be discussed in the next [post](**LINK%20TO%20POST2**)). The binary function is *binary* because it returns only two value which are often 1 and 0. If the opportunity is reachable from to within some sort of travel cost threshold , it returns a 1 for that trip. Conversely, it returns 0 if the travel cost is above a certain threshold , meaning the opportunity exceeds the cost that people are willing to travel to reach it.

Threshold should be selected carefully to reflect the observed or assumed travel behavior for the situation of interest. For instance, assume the travel cost is in the units of car travel minutes. Does only counting the potential interaction opportunities for the population in a region accessing destinations within a 0 to 15 minute range (the travel cost) make sense for the context of accessibility analysis? If yes, then the threshold in this function would be 15. This means that the opportunities that can be reached by the population will only be counted if those opportunities can be reached within 15 minutes, if not, they are not counted at all (assigned a value of 0) in the accessibility measure ([Equation 1](#eq-hansen-access)).

Three more commonly used forms of impedance functions in accessibility analysis are important to consider: these functions should be interpreted differently than the binary function ([Equation 2](#eq-binary-access)) as they are all theoretical [*probability density functions*](https://en.wikipedia.org/wiki/Probability_density_function) (PDF). In conceptualizing the impedance function as forms of PDF, the values can be interpreted as the *probability density* of a trip occurring for each value of . If probability values are plotted on the y-axis for each travel cost along the x-axis, the probability of a trip occurring between a certain range of is the area under the curve. Important to note is that the area under a PDF always sums to 1, i.e., 100% probability that the trip between the minimum and maximum will occur.

The **uniform distribution** PDF looks very similar to the binary function (shown in ([Equation 2](#eq-binary-access))), as it only returns one of two values. However, it also has the property of PDFs - the area under the curve for the range of is always 1 (i,e., 100% probability that the trip between the minimum and maximum will occur). The general form for the uniform distribution PDF is shown in ([Equation 3](#eq-uniform-imped)). The parameters that the analyst chooses are and : these represent the maximum and minimum travel costs (i.e., the range) that describe the observed or assumed willingness to reach destinations. If the trip is of a travel cost that is within this range, it returns a value of ). Outside of this range, the function can be interpreted to assume that the potential for interaction is zero so the function returns a 0 for trips of those travel costs.

The next two functions are the **exponential distribution** and the **gamma distribution** (part of the exponential function family but utilize the gamma function ). The theoretical form of these two PDFs are shown in ([Equation 4](#eq-exp-imped)) and ([Equation 5](#eq-gamma-imped)). The analyst must select parameters for these functions represented by (exponential) and and (gamma).

For the **exponential distribution**, the probability of a trip occurring is always highest at the lowest value of travel cost (e.g., a trip that has a travel cost of 1 has a higher probability density than a trip with a travel cost of 10). The mediates the rate of the exponential curve; specifically, the higher the parameter value, the higher the rate of travel cost decay. So at a value that is large, the majority of trips occur within a smaller range than if the was a smaller value. Though the exponential distribution is more complex than the uniform, it allows the analyst to model travel behaviour without having to select binary cut-off travel cost beyond which opportunities are no longer counted (like in the binary function or uniform distribution PDF).

Binary travel cost thresholds may not make sense in some applications: for instance, is it true that in a hypothetical example, no trips occur beyond 15 travel cost units for a region, population and opportunities of interest? Is this a fair assumption to make about the travel behaviour? Maybe it is more fair to say that the probability of a trip occurring decreases as the travel cost increases in an exponentially decaying way (as informed by the parameter). In this case, it would be worth while considering the exponential distribution instead of the uniform distribution.

Unlike the exponential distribution, the probability of a trip occurring is not always highest at the lowest value of travel cost for the **gamma distribution**. In fact, for the **gamma distribution** the probability is often low at the low costs, higher at mid-costs, and low again at high costs. The and parameters controls the rate and shape of the gamma curve, like controls the rate of the exponential curve. The higher the (gamma rate) parameter, the higher the probability of the majority of trips occurring within a low travel cost range. So at low (gamma rate) parameter values, the same probability is spread across a larger range of travel costs. For the (shape) parameter, the higher value, the higher the probability density of trips with a higher mean travel cost.

Values for both and are used in the gamma distribution, so it is more complex in formulation than the exponential. However, the gamma may be more useful in modelling specific travel behaviour. Namely, if the population’s travel behaviour is less likely to occur at short travel times, more likely at mid-range travel times, and less likely at long travel times, the gamma distribution can be calibrated to match this pattern. For instance, this form of travel behaviour can occur within observed home-to-work commutes from predominately single-use zoned regions: trips are less likely to occur at short travel times for a region (as a result of single-use residential zoning), are more likely at mid-range travel costs (commuting to a central business district), and less likely at long travel costs (few super-commuters). Representing this travel behaviour pattern cannot be accurately captured by the exponential distribution as short travel times *should* have a low values of . Additionally, the use of the uniform distribution may be inaccurate in this situation as it requires the analyst to select min. and max. travel cost thresholds such that the opportunities that short- and long- travelling population potentially interactions is not counted (i.e., returns value of 0). The analyst must ask themselves if it make sense to not count the potential opportunities that can be reached by the *few*, but still occurring, short- and long- travel cost trips.

In summary, these three discussed PDFs can take a value of probability from 0 to approaching infinity for all positive , where the range depends on the analyst-defined parameters. They are presented in order of increasingly complexity, but as the complexity increases, the flexibility of explaining the travel behaviour also increases. For convenience, I created an interactive R Shiny Application with these three distribution PDFs [here](https://soukhova.shinyapps.io/Impedance-explained-shiny-app/). Feel free to experiment with the parameter values and conceptualize what each function may be assuming about travel behaviour by interpreting the “probability density of trip” (y-axis) at values of travel cost (x-axis) .

# An empirical example: calibrating a model that reflects the impedance of travel behaviour

In all the impedance function forms presented, the analyst must define the parameters. A useful technique to calibrate the parameters is by using empirically observed origin-to-destination (OD) travel data. This empirical data can be used to build a trip length distribution (TLD). The TLD is the *empirical* PDF of the travel costs associated with the OD trips. In other words, this distribution reflects observed travel patterns: we can use the TLD to tell us how likely an observed trip of a certain travel cost is to occur for the population and region of interest. Based on this TLD, we can select the best fitting theoretical PDF form (e.g., uniform, exponential, gamma), fit the associated parameters (e.g., & $T\_{max$, , or & ) and use the calibrated theoretical PDF to carry the assumptions about travel behaviour into the accessibility calculation.

Below I demonstrate an overview of calibrating a PDF for a sample of empirical home-to-work flows taken from workers who live and work within Hamilton Center from the R data package [{TTS2016R}](https://soukhova.github.io/TTS2016R/) (Soukhov and Paez 2022). The flows are taken at the spatial unit of traffic analysis zones (TAZ). This package contains a subset of home-to-work flows the 2016 Transportation Tomorrow Survey (TTS) as well as calculated road-network car travel times (calculated using [{r5r}](https://doi.org/10.32866/001c.21262) (Rafael H. M. Pereira et al. 2021)). {TTS2016R} is detailed in this publication (Soukhov and Páez 2023) and is freely available [here](https://soukhova.github.io/TTS2016R/).

The TLD for this empirical data is shown in black in the plot below ([Figure 1](#fig-TLD-empirical)):

|  |
| --- |
| Figure 1: Trip length distribution of home-to-work trips (in estimated minutes by car) for Hamilton Center. |

As explained, the TLD is the *empirical* PDF of the travel costs associated with the OD trips. Like all PDFs, the y-axis represents the probability density of the value at the x-axis of occurring. In our example, is the probability density of a trip at a given travel cost in minutes of travel, . It can be observed that the probability density of travelling is highest when travel time is around 11 minutes. It can also be seen that around the 20 min mark, the probability density levels off. In other words, the probability of a trip of length 0 to 20 mins occurring is 100%. Trips outside of this range make up the remaining probability.

Now let’s fit the parameters of the uniform, exponential, and gamma functions ([Equation 3](#eq-uniform-imped), [Equation 4](#eq-exp-imped), [Equation 5](#eq-gamma-imped)) as closely to the TLD captured in [Figure 1](#fig-TLD-empirical) above. The R package {{fitdistrplus}](https://cloud.r-project.org/web/packages/fitdistrplus/index.html) was used to generate parameters that best-fit the TLD: the moments matching estimation (MME) fitting-method and the Nelder-Mead direct optimization algorithms are used (Delignette-Muller and Dutang 2015). The default values for all three functions are summarized in the following, but if interest in the replicating and reproducing these results, see the R code in the Quatro document within this [GitHub repository](https://github.com/soukhova/MJ-Accessibility-Blogs/Impedance-explained.qmd).

* Uniform function ( - red): and is 0 and 19 mins, respectively.
* Exponential function ( - green)\*: (rate) is 0.1
* Gamma function ( - blue)\*: (shape) is 3 and (rate) is 0.4

|  |
| --- |
| Figure 2: Trip length distribution (empirical) with fitted theoretical PDFs of home-to-work trips (in estimated minutes by car) for Hamilton Center. |

Functions in [Figure 2](#fig-TLD-all) can be interpreted in the following way: the higher the , the higher the probability density of travelling to reach the opportunities at the destination.[[1]](#footnote-40)

The uniform impedance function (red), when implemented into an accessibility calculation, would assume that the population is indifferent to *changes* in travel cost. The population at an origin is assumed to either totally interact with an opportunity (if it’s a trip between 0 to 19 minutes - the thresholds) or not interact at all.

If the above exponential or gamma function was implemented in an accessibility calculation, then the analyst is assuming the population is much more sensitive to changes in travel cost. However, the exponential and gamma functions are quite a different shape so they depict a different response to the probability of traveling given a travel cost .

The exponential function (green) is more intuitive to understand: the shorter the travel cost , the higher the value. However, when compared to the empirical TLD (black) curve, we can see that the observed travel behaviour does not closely match this curve. Trip lengths that are 11 mins in length have the highest probability density of occurring and trips that are longer and shorter than this length occur less often and are assigned decreasing values. For these reasons, the gamma function (blue) provides a fit that is closest to the empirical curve at the cost of a more complex mathematical formulation (see [Equation 5](#eq-gamma-imped)).

The impedance function reflects significant assumptions about travel behaviour. The selection of the type of function and associated parameters reflects how much impedance the modeled population faces reaching opportunities and hence their potential interaction. How the impedance function is used to explain accessibility functions will be discussed in next [post](**LINK%20TO%20POST2**). Again, feel free to explore the parameters interactively for the uniform, exponential and gamma distributions using the interactive Shiny R Application [here](https://soukhova.shinyapps.io/Impedance-explained-shiny-app/).

*The TLD used in this post is a subset of data from {TTS2016R}, the goodness-of-fit criteria and diagnostics from {fitdistrplus} are used for model parameter selection, plots are generated using {ggplot2}, and spatial objects are manipulated using {sf}, along with base {R} functions. Feel free to view all the code and text in this post (including the interactive plot) in the* [*GitHub repository*](https://github.com/soukhova/MJ-Accessibility-Blogs)

## References

Delignette-Muller, Marie Laure, and Christophe Dutang. 2015. “fitdistrplus: An R Package for Fitting Distributions.” *Journal of Statistical Software* 64 (4): 1–34. <https://doi.org/10.18637/jss.v064.i04>.

Hansen, Walter G. 1959. “How Accessibility Shapes Land Use.” *Journal of the American Institute of Planners* 25 (2): 73–76. <https://doi.org/10.1080/01944365908978307>.

Levinson, David, and David King. 2020. *Transport Access Manual: A Guide for Measuring Connection Between People and Places.* University of Sydney. <https://ses.library.usyd.edu.au/handle/2123/23733>.

Pereira, Rafael H. M., and Daniel Herszenhut. 2023. *Introduction to Urban Accessibility: A Practical Guide with r*. <https://doi.org/10.38116/9786556350547>.

Rafael H. M. Pereira, Marcus Saraiva, Daniel Herszenhut, Carlos Kaue Vieira Braga, and Matthew Wigginton Conway. 2021. “R5r: Rapid Realistic Routing on Multimodal Transport Networks with R5 in r.” *Findings*. <https://doi.org/10.32866/001c.21262>.

Soukhov, Anastasia, and Antonio Paez. 2022. *TTS2016R: An Augmented 2016 Transportation Tomorrow Survey (TTS) Data Package: Worker and Place of Employment Counts, Trips and Estimated Travel Time to Work in the Greater Golden Horsehoe Area, Canada*. <https://soukhova.github.io/TTS2016R/>.

Soukhov, Anastasia, and Antonio Páez. 2023. “TTS2016R: A Data Set to Study Population and Employment Patterns from the 2016 Transportation Tomorrow Survey in the Greater Golden Horseshoe Area, Ontario, Canada.” *Environment and Planning B: Urban Analytics and City Science*, January, 23998083221146781. <https://doi.org/10.1177/23998083221146781>.

Stewart, John Q. 1948. “Demographic Gravitation: Evidence and Applications.” *Sociometry* 11 (1/2): 31–58. <https://doi.org/10.2307/2785468>.

1. Using {fitdistrplus}, the parameters in all theoretical functions were selected through an optimization algorithm that minimizes the differences between all possible parameter range(s) and the empirical function for each theoretical function. [↑](#footnote-ref-40)