

**technical notes for the canopy method**

Drafted by Nicholas mader (nmader@chapinhall.org)

# Introduction

CANOPY--a now-general name adapted based on its original application: Chicago Allocation to Neighborhood-Oriented Programs for Youth--is effectively a type of recommendation engine that can help city planners make decisions about how to allocate resources across neighborhoods, to best meet the goals that they specify.

CANOPY is a software-based assistant for helping make city planning decisions, much in the same way that Google Maps helps plan travel decisions. Google Maps is useful because it can account for many simultaneous considerations—more than what a human can easily track down and integrate—of routes, alternatives, speed limits, traffic patterns when making suggestions. Because Google Maps works quickly, it is also helpful by offering multiple different options, and being easy to make recommendations based on scenarios. In the same way, CANOPY is intended to make city planning decisions smarter but also simpler, by simultaneously accounting for spatial patterns of where households vs. public facilities are located, neighborhood data, and patterns of how households make decisions about whether and where to access public facilities. CANOPY receives input about the goals of decision-makers, and makes suggestions without taking over any of human thinking and discussion that need to be made to select a course of action.

This document formalizes the general statistical and mathematical model which underlies CANOPY. The document “Introduction to CANOPY - Chicago Allocation to Neighborhood-Oriented Programs for Youth Algorithm - With Maps” gives a non-technical introduction to the same ideas.

## agent choice model

Suppose that there are decision-making “agents”, indexed by , who each make a selection among alternatives, indexed by . We also allow for agents to make no selection at all, represented by , such that the expanded set of alternatives are mutually exclusive. A real-world example is where many households with a child eligible for Head Start have a choice of potentially many centers where to enroll their child, but may also realistically not make a choice to enroll their child anywhere.

Let represent a vector of characteristics that is known about each agent, a vector of characteristics known about a given alternative, and represent a vector which combines agent and alternative data. Continuing the example, may represent the address, race/ethnicity, household composition, and education of parents; may represent whether the provider is school-, center- or home-based, its staff-to-child ratio, its address, and summaries of the types of households that previously enrolled there; and may reflect information such as the distance between and .

Let represent the latent valuation that agent has for alternative , and let it be expressed by:

where we pose the functional form that value is a linear combination of all factors included in , and where the component of value reflect all salient components of valuation that are not known (or “observed” in a data set) in analysis of agent decision-making. To set the “location” of this series of latent values, we normalize the scale such that .

Let represent the index value of the alternative chosen by agent . We presume that agent selects alternative if and only if . We can use this definition for and its presumed functional form to connect agent choice of alternative to observed factors. This mapping is only probabilistic, given that conditioning on only observed factors will generally result in a range of chosen alternatives given the distribution of :

Assuming that the terms follow a Type II Extreme Value distribution, the follow functional form for probabilistic choice selection for each agent obtains:

For simplicity in the following discussion, we will use the notational simplification: .

## City planner’s problem

Our current working representation of the city planner problem is to allocate resources in such a way as to optimize an objective based on which choices are made by which types of agents. We expand our functional form of the choice model above to allow for this resource allocation:

which reflects the addition of to the valuation . In this setup, the policy action available to city planners is to determine the allocation of resources to each alternative—notated as the collection , or vector —such that where represents the total amount of resources that can be distributed among alternatives.

The city planner’s problem can thus be expressed as:

Because is a nonlinear function of , this represents a nonlinear constrained optimization problem.

Choice of the weights can be used to represent a wide range of policy goals. A scheme where:

would reflect a policy goal where planners aim to have as many agents select non-0 alternatives. This reflects policy exercises where all agents are evenly targeted, and where all alternatives are equally valuable. Examples may be get as many individuals to obtain flu shots or, among agents who are eligible for a public support program, to come to a service center to be enrolled.

Alternatively, policy weights could be chosen to vary with agent characteristics, representing a particular targeting of agents within the population:

For example, children are eligible to enroll in pre-Kindergarten through Head Start if their family is below the poverty line. However, policy makers may wish to place a premium on successfully targeting children who are particularly disadvantaged, such as

Or, the policy weights could be chosen to reflect a combination of targeting certain agent subpopulations and certain alternatives:

such as:

where:

* there is no value to children not being enrolled anywhere (i.e. when ), is the distance between and so that prioritizes choices that are close to home;
* is the ratio of ’s family income to the Federal Poverty Line, so that the term smoothly prioritizes youth in proportion to depth of poverty circumstances;
* is an indicator of whether is a center- (versus home-)based provider, so that policy makers generally favor enrollment in center-based locations;
* is an indicator of whether represents a housing with a working single parent, and is an indicator of whether alternative provides late hours that accommodate work schedules; and
* each term is a weight that is given to the corresponding term of the function.

Clearly, this example shows how many components of policy maker value can be embedded into this framework, to tailor the solution to policy goals. Even in this one example, a wide range of priorities could be set by adjusting the set of values to put different priority weights on the goals represented in each term.

## heuristic description of the simulated annealing optimizer

Because of the structure of the terms, the planner’s problem has no analytical solution. And because the objective function in most applications will be non-convex in , gradient descent methods will not reliably obtain a global optimum to solve the problem. For these reasons, we have chosen simulated annealing (SA) as a global optimization heuristic, which attempts to explore the parameter space broadly early in its run (to avoid settling into optima local to the initialization), and focuses on local convergence later in the run when the space has been explored. The SA procedure requires the objective function , an initial parameter allocation, a (decreasing) temperature function which is purely a function of the iteration currently taken, a transition function which specifies how neighboring states are drawn. Its steps are as follows:

1. From the current allocation state , use the transition function to draw a proposed allocation state ;
2. Compare the value of the objective function for with that for :
   1. Accept as the new allocation state with probability:
3. If the maximum number of iterations has been reached (which is the only termination condition), then stop. Otherwise, whether or not was accepted, update the function and return to Step 1.

Note, in Step 2, is accepted with certainty if it improves the objective function. The SA procedure may also accept even if it disimproves the objective function, since it may lead to a better area of the state space which will lead to a better optimum. However, the probability that a dis-improving will be accepted decreases in proportion to the extent that it dis-improves the objective, reflecting the willingness to experiment with it. That probability is increasing in the temperature function , which itself decreases as a procedure goes on, reflecting the fact that experimentation is more likely at early stages, and less likely at later ones.

In our implementation, we follow the conventions that is monotonically decreasing, where is the final iteration, and . Because the SA routine does not guarantee monotonic progress in the objective function through its run, it is possible that the with the largest value of the objective function is not . For that reason, our SA routine saves the allocation with the largest objective value that has been seen, and reports that at the end of its run.

## sample application used for prototyping canopy

The application chosen for the prototype uses public sources of data to avoid the process of obtaining sensitive data, and reflects a contrived policy problem to showcase CANOPY without involving the distractions of a high-stakes policy example.

The sample policy is one where the city is attempting to maximize youth participation in neighborhood basketball programs, and has a given number of staff which it can allocate among the city’s basketball courts to strategically encourage youth to attend. They positively value having any youth attend, but put more weight on reaching youth who are particularly disadvantaged from a family income standpoint.

**Choice Model.** Youth value participation in basketball programs, but their value for participating at a given site decreases with the distance that they would have to travel, and the more crime there is in the youth’s neighborhood, the less willing they are to travel any given increment of distance (i.e., the rate that distance decreases valuation is decreasing in neighborhood crime). The youth choice model is calibrated as:

where is distance between youth and court , is crime in the youth’s neighborhood, is the number of staff allocated to the court, and is the Type II Extreme Value error (with mean zero and variance of 1).

Note, of course, that in a real-world application, the coefficient parameters in the choice function would be unknown, and would be inferred from observed patterns of using discrete choice regression models (such as the conditional logit model used to set up the agent choice structure)[[1]](#footnote-1)

**Policy Weights.** The policy weights chosen for the policy makers are:

where is the ratio of youth ’s family income to the Federal Poverty Line.

**Youth Data.** Data from the 2008-2012 American Community Survey (ACS) 5-year release are used to get age-by-poverty counts at the tract level. We use data from table B17024[[2]](#footnote-2) to identify the number of youth aged 12 to 17 years old, with family incomes fit into the categorical bins above, living in each census tract in the city of Chicago, IL. Youth are initially assigned a residential location equal to the centroid of the census tract of residence, and then the latitude and longitude are “jittered” in either direction by adding a uniformly-distributed number between -1/2 and +1/2 of a mile.

**Crime Data.** Violent crimes per 100 residents in each census tract are calculated by taking neighborhood crime data are taken from the city of Chicago data portal (<https://data.cityofchicago.org/>) for the year 2011, determining whether the crime is a violent one using the Uniform Crime Reporting codes employed by the Chicago Police Department (<http://gis.chicagopolice.org/clearmap_crime_sums/crime_types.html>), and using a denominator of all residents of each census tract taken from the 2007-2011 ACS 5-year release.[[3]](#footnote-3)

**Court and Distance Data.** Data on the locations of outdoor basketball courts was taken from the Chicago Park District website (<http://www.chicagoparkdistrict.com/facilities/basketball-courts-outdoor/>). These locations were geocoded using the Google Maps API. The distance between each youth and basketball court pair was approximated as a “city block” distance (or L1 norm), i.e. .

**Allocation.** For simplicity, the city was endowed with enough staff to be able to assign exactly one per court. The youth choice model was calibrated so that a uniform (one staff per court) allocation would (on top of other spatial patterns of crime, residence, and court locations) yield a low, but positive, probability that the average youth would participate in the basketball program. The hope was that this uniform initial distribution of staff resources would be an easy point of comparison with CANOPY-recommended allocations which would recommend leaving certain courts with no staff, and others with potentially many.

**Limitations in Human Planning Processes.** A human planning process unassisted by CANOPY could make use of the same information available to CANOPY—in the form of maps of youth/court locations, poverty, and crime—but have a more difficult time integrating those data. For each allocation decision, a calculation would need to be made to determine which among all courts would be benefitted by gaining a given staff member, taking into account all residential, crime, and poverty information; the numbers of staff already allocated to nearby courts; and the ways that all of these factors combine into a probability that various youth would now be more likely to become engaged.

While it is not feasible for human decision makers to keep and use all of this information in mind, the use of certain heuristics make unassisted human planning process tractable and may yield reasonable policy success. For this reason, the CANOPY application will compare the expected success of its recommended allocation with those resulting from rule-of-thumb solutions. Examples of rule-of-thumb solutions are (1) allocating all staff evenly among courts; (2) allocating staff in proportion to population density in the area; (3) allocating staff in proportion to neighborhood poverty; or (4) allocating staff in proportion to an index constructed of the number of nearby (e.g. within a certain radius) youth, weighted by poverty and/or crime circumstances.

Note that the more complex a planning process becomes—such as many additional factors that play into the agent choice model, more characteristics of alternatives, and the more nuanced the policy goals—the greater the advantage that CANOPY will have over rule-of-thumb planning solutions.

**Benefits of CANOPY.** By contrast to the human planning process, CANOPY offers key benefits in terms of attention to detail and speed. Planning processes involving CANOPY use computer-based assistance for the tasks that are hardest for humans—keeping in mind and working with detailed data on agent decision processes, choice context, and detailed policy goals—while accommodating aspects of the planning process that cannot be done without human consensus building and judgment—i.e. determining and articulating the policy goals, and making final decisions on what allocation to make. Because use of CANOPY enhances human policy planning, it will yield better policy outcomes by making more complete use of the information that is available.

CANOPY’s speed—where the algorithm can try roughly thousands of different candidate policy solutions per second—means that it will rapidly produce policy recommendations in a matter of minutes instead of days or weeks of human time. This speed also means that planners can use CANOPY to run and compare many different scenarios, varying how priorities are set, or what goals could be reached with different levels of starting resources.

Other benefits:

1. Consistent decision-making – CANOPY is guaranteed to use the same precise policy guidance when making the first allocation decision as the last. By contrast, human decision-making can be
2. Necessitates a clear statement regarding planning goals

1. See <http://en.wikipedia.org/wiki/Discrete_choice#F._Logit_with_variables_that_vary_over_alternatives_.28also_called_conditional_logit.29>. [↑](#footnote-ref-1)
2. See <http://www.socialexplorer.com/data/ACS2012/metadata/?ds=American+Community+Survey+2012&table=B17024> for reference to this ACS table. [↑](#footnote-ref-2)
3. These 2011 dates are somewhat out of line with the 2012 data on youth locations. This was less a matter of choice than convenience in bringing previously-develop (by us) data sources. As a somewhat lower priority, this may be better aligned in the future. [↑](#footnote-ref-3)