Project List 2

1. Suppose you are a shepherd. The sheep you raise have a lifespan of several years. Lambs are born in a particular season each year, and after this season you may sell some of your flock. You are capable of caring for up to N adult sheep, and will always sell excess sheep if the flock exceeds that size after the lambs are born. You may assume that the birth and death rates (per sheep per year) do not depend on the size of the flock, but they do depend on the age of the sheep. Relevant data can be found in: G. Caughley, "Parameters for seasonally breeding populations", Ecology **48** (1967), 834–839. Suppose that when you sell part of your flock, you sell the same fraction of each age group. How many sheep can you expect to sell each year on a continuing basis? What if you only sell sheep of selected ages; can you sell more sheep per year, and if so what is the best strategy? If only young sheep are marketable, how will this affect your yearly sales?

2. In many situations data consisting of a list of items is stored in an "unsorted" fashion, so that it must be searched sequentially (first item, second item, ...) to find a desired item. This may be computer data, or a stack of files on one's desk, for example Suppose each item has a certain probability of being searched for in a given instance, and the probabilities are known. What is the average search time, and how should the items be ordered to minimize this time?

   Next, suppose that the probabilities are unknown (or in practice they might be changing slowly over time). Consider various "self-organizing" strategies to try to reduce the average search time by changing the order of the list after each search. For example, each time there is a search, the item found might be moved to the front of the list (this is a common strategy in the desk scenario). Or, the item found might be moved partway to the front, or switch places with another item. How will the positions of the items tend to evolve, and what will the average search time be in the long run (after the list has had a lot of time to reorganize itself).

   Another variation on this problem is when there are different categories of data storage with different access times — for example, a desktop *vs.* a filing cabinet *vs.* a storage room, or the memory cache on a computer chip *vs.* standard memory *vs.* a hard disk. Generally the categories with faster access times are smaller and those with slower access times. The access times may be sufficiently different that (unlike above) one can regard the access time of an item in a given category as being independent of its location within that category. How might "self-organizing" strategies work in this situation?

3. A supermarket must decide how many cashiers to employ at various times of day (and night). The rate of incoming customers is fairly predictable but varies substantially by time of day. However, as the number of customers waiting in line gets larger, the proportion of incoming customers who turn around and go to another store also grows. How should the supermarket design the work schedule for its cashiers?

   Consider also the situation of a company that takes sales orders by telephone – how many operators should the company employ at different times? Again assume that the rate of incoming calls is fairly predictable, and in this situation that the company can lose business by having customers hang up after waiting "on hold" for too long. How would your model differ in this case?

4. Workers in a warehouse fill orders by attaching each written order to a container and moving the container along a series of aisles, filling the container with items listed on the order. Having each order filled by a single worker was found to be inefficient because the workers got in each others' way, so instead each aisle is divided up into equal sized sections, and a worker is assigned to each section. Each worker takes a container from the start of his/her section, puts the ordered items from the section into the container, leaves the container at the end of the section for the next worker to pick up, and returns to the start for the next container. Each container passes through the sections in the same order, so that there is a "first worker" who starts each order and a "last worker" who completes each order. (Picture for instance a warehouse for CDs, where the orders and the inventory are alphabetized. The first worked might be assigned to letters A and B, the next to letters C–E, and son on.) There still inefficiencies in this scheme though, because different orders take different amounts of time at each section, and different workers work at different speeds. Thus sometimes a worker must wait a while because either there are no ready containers from the previous worker, or there are too many containers waiting for the next worker.

   The relative speeds at which the workers fill orders is measured, and three modifications are proposed to improve efficiency using the same group of workers. One is to resize the sections to be proportional to the speeds of each of the workers. Another is to slightly decrease the number of sections and assign the displaced workers to be "floaters" who can temporarily help a worker who is falling behind, and can also fill in for a worker who is taking a break. Another is to eliminate the section boundaries and instead assign workers to keep themselves in a particular order but to continue to fill a container until the next worker is available to take the container, then go back to the previous worker and take his/her container, and so on. Some of these modifications could be combined as well. Compare the different possible schemes. Does one scheme seem better than the others or does it depend on the parameters? In each case, how does the ordering of the workers (slowest to fastest, fastest to slowest, or some other combination) affect the overall performance?

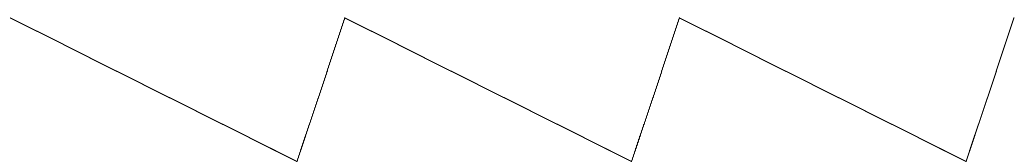
5. Virtually any medical test for a disorder has a chance of both "false positives" (people who don't have the disorder testing positive) and "false negatives" (people who do have the disorder testing negative). Suppose there is a test for a certain disorder that is fairly accurate but is expensive. Treatment for the disorder costs less the earlier the disorder is diagnosed, but in any case is substantially more expensive than the test. From the point of view of a health care provider that must incur the costs, should the test be given to everyone as part of a regular checkup, or should it be given only to people who show symptoms of the disorder? How does the answer depend on the relative costs, the incidence of the disorder in the population, the typical time between contracting the disorder and showing symptoms, *etc.*?

   Suppose that someone discovers a new test that is substantially cheaper than the present test but is also significantly less accurate. When should the new test be given, and should the present test still be used in some circumstances?

There are a lot of parameters involved here, and the goal is to identify the relevant parameters and come up with a general procedure for deciding (from an economic point of view at least) what testing strategy to use given the values of the parameters. But you can check the web page `http://www.aap.org/policy/01565.html` to get an idea of the actual values of parameters like incidence, testing cost, and false positive and negative rates for various pediatric disorders and tests.

6. Suppose we have a mixture of two types of particles of different masses and/or sizes and we want to separate the two. One possibility is to put them in a slightly inclined, corrugated tube and shake the tube. To make the geometry a little simpler, let's think of a corrugated incline plane rather than a tube. Construct a model that describes how a group of particles of a particular mass and size moves downhill, and use it to discuss how, given the mass and size of two different types of particles, one should choose the parameters such as the size of the corrugation, the angle of incline, and the amount of shaking to best separate an initial group of mixed particles at roughly the same height. Assume that the particles are spread out enough that you do not need to worry about them interacting with each other. The idea here is not to derive a precise model from the fundamental laws of physics, but rather to come up with a reasonable phenomenological model based on physical intuition.

   Bonus question: suppose the corrugation, rather than a symmetrical sine wave or sawtooth shape, is shaped more like a ratchet:

   Do you think it is possible that one type of particle would tend to move uphill whereas another type would tend to move downhill? How might this be advantageous over the case of differing speeds of downhill propagation?

7. Modern computer chips generate a good deal of heat, and to avoid damaging themselves they require some external apparatus to dissipate the heat. In addition to a fan for ventilation, manufacturers often attach a "heat sink" to the back of the chip. The heat sink is made of some material that conducts heat well, and is typically shaped like a vertically elongated version of the surface of a waffle iron; in other words, a rectangular grid of spikes or "pins". Why is it shaped like that? (For that matter, why is a waffle iron shaped the way it is?) Assume that the computer chip is a flat square generating heat uniformly across its surface. What should be the dimensions of the pins that make up the heat sink, and how far apart should they be spaced, in order to keep the chip as cool as possible?

   The web page `http://www.indek.com/products/DCpator/hs_data.htm` gives one manufacturer's specifications for various sizes of heat sinks. What relationship(s) between the size of the heat sink, size and shape of the pins, and spacing of the pins does this data suggest and can you justify the choices made by this manufacturer with your model?

8. Instructions for cooking food in boiling water or in a conventional oven generally direct that the water be brought to a boil or the oven be preheated before the food is inserted. How much sooner would the food be cooked if inserted at the same time the heat is turned on? What drawbacks might there be to cooking the food in this manner, either from the point of view of the person eating the food or the person writing the instructions? Suppose you work for a company that sells packaged food, and the marketing department wants to give alternate cooking instructions for "people on the go" who do not want to wait for the water to boil or the oven to preheat. How would you formulate such instructions and what would your advice be? How do your answers to these questions depend on the cooking method (boiling water *vs.* oven) and the type of food?

9. Traffic engineers find that on a highway, the maximum flow rate (number of vehicles passing a given point per unit time) is generally achieved at speeds of 30 to 40 mph. This is due to a trade-off between the density of vehicles on the highway and the speed at which they travel, which decreases with density. Develop a model that describes how individual vehicles in a given highway lane move, assuming that the speed a driver chooses is based primarily on the speed of and distance to the vehicle immediately ahead of the driver. Use this model to relate traffic speed to density and try to reproduce the result describe above. Now add to your model the effect of a traffic light which is alternately green for a certain time period T_1 and red for a certain time period T_2. What is the average flow rate of vehicles through the intersection as a function of T_1, T_2, and the flow rate of vehicles approaching the intersection, and under what circumstances does traffic the number of vehicles waiting at the light grow with each cycle? If two highways cross at the intersection and you know for each highway the flow rate of approaching vehicles, what values of T_1 and T_2 maximize the combined flow rate of vehicles through the intersection? Place a cap on the cycle time $T_1 + T_2$ if necessary. Is it possible that the optimum time allotment results in traffic backing up on one highway and not the other?

10. This project concerns both some simple methods for quantitative prediction and how to evaluate the performance of different methods, using a particular example of weather. First, develop a means for comparing how accurate different weather forecasts are over time, and use it to compare different local weather forecasts over the coming month. Many sources now give 5-day forecasts, for example:

    `http://www.washingtonpost.com/wp-srv/weather/`
    `http://www.wjla.com/weather/`
    `http://www.weather.com/weather/local/USDC0001`

    How much more accurate are the forecasts for tomorrow's weather than those for the day after that, and for the next day, and so on?

    Second, make your own predictions based on simple linear models, along the following lines. One of the easiest predictions you can make about tomorrow's high temperature is that it should be somewhere near today's high temperature. Usually this is a decent prediction, but probably less accurate than a meteorologist's forecast! You can make the predictions a little more sophisticated by using historical data and the method of least squares to determine the "best fit" value of a and b for the prediction

$$T_{n+1}^{(p)} = aT_n + b,$$

where $T_{n+1}^{(p)}$ represents the prediction for tomorrow's high temperature and T_n represents today's high temperature. One possible next step is to predict tomorrow's high temperature in terms of today's and yesterday's; again you can use historical data to determine the "best" constants in a model of the form

$$T_{n+1}^{(p)} = aT_n + bT_{n-1} + c,$$

Theoretically you can continue to improve the prediction by adding more variables to your model, in the form of temperatures several days in the past, but you should soon reach a point of diminishing returns. You can use the diagnostic you developed in the first part of this project to compare your linear predictors of different orders, and see how many days in the past you need to take into account before you stop seeing any significant improvement. Other variables you might take into account to predict tomorrow's high temperature are the historical average temperature for tomorrow's date, today's low temperature, etc. These may be more relevant than the temperature several days ago. See how well you can do and compare your predictions with the meteorologists' forecasts.

11. How likely is it that the best team wins the NCAA basketball tournament? Develop a model that describes a typical distribution of skill among the 64 teams in an NCAA basketball tournament and the probability of the better team winning a game as a function of the skill levels of the two teams playing. Take into account scores of past games in developing the model, and demonstrate that your model fits past results reasonably well. Does the same model fit both the men's and women's tournaments? Simulate a large number of tournaments using your model and give probabilities for teams of different skill levels to win the tournament. Again, compare your results to actual results.

12. Formulate your own project!