

Some Suggestions for Projects (Group or Individual)

I would like to encourage you to consider doing a project, either by yourself individually, or as part of a group effort. The ground rules are very simple. You should (1) go at least one step beyond what we have done in class, and (2) you should prove that you have made the subject your own. Part (1) here means that you should do something recognizably different from anything done already or to be done in the course. Part (2) means that you don't just go to a reference and copy out something new you find there. One good way to show that you have "made a subject your own" is to give examples and analyze them or do selected exercises about a new topic. The papers should be at least five pages long, longer if they represent a group effort (that should be arranged with me beforehand) or if you are taking the writing supplement.

As I have mentioned in class already, there are *many* possible topics for this, although I will first repeat the suggestion that if you are studying an area in which probability is applied, one very useful thing you could do for the project is to delve into one significant application of probability in the field of your interest. If you don't already know such an application, *use this as an excuse to talk with one of the professors in your field*, and ask him or her to direct you to such an application. Not all of these applications will be accessible to you yet, since there are a lot of things about probability we still haven't covered. But that is one necessary component of a project: go at least one step beyond what we have been doing in class! I propose a few topics below which are either directly accessible to you now, or will be so. I repeat, this list is suggestive, and is **not** meant to be exclusive.

Remember, your proposed topics are due Friday, March 7, so I can approve them before the second hourly exam. You should turn in a brief note with the topic, one or two suggested references and a name of a faculty person you have contacted about this, if that seems relevant. If you are interested in doing a group project with other students, you should let me know the group at that time as well. To find a possible group, discuss this possibility with your group homework group or with those on the class address list who have similar interests (major). Finally, I have stacks of old examples of this kind of project which you are free to have a look at for a more concrete idea of what is expected.

1. Statistics. There are, naturally, many things one could do here. This would be a good opportunity to look into the possible topics that Stat 426, one of the follow-up courses to this one, will treat. Put more directly, given a large sample of data, how does one use probability to analyze it. Even more directly, given a list of H's and T's, how can you tell whether it is more likely to be from a biased coin or a fair coin whose data is in a state of fluctuation from the expected values? Or, again, what is the chi-squared distribution and how is it used to measure the distribution of errors of measurement, or dispersion in the values assumed by an **RV**? Or, what are the concepts of sample random variable, point estimation, interval estimation, maximal likelihood estimation and least squares

estimation of probabilistic parameters (like p and λ), or statistical testing of hypotheses?

2. Finance and economics. How does one use probability as a means to model, understand, estimate and prepare for risk, either in settings such as life and other forms of insurance, or in financial markets, especially in the estimation and pricing of derivative securities (futures and options, e.g.). These are areas of great recent interest and optimistic job prospects, at least for now. (There, the element of risk enters already!)

More specifically, what does random walk have to do with the stock market? What is the “lognormal hypothesis” about stocks? What is the Black-Scholes model for pricing of options?

3. Random Walks. I have already mentioned this above with finance in mind, but there are myriad other examples. How is this related to the normal distribution? (We will begin discussing this in class.) How is the normal distribution related to heat diffusion? (It is!) And finally, an example which I find very interesting, how is probability used to model and understand “genetic drift”? We are already doing some very elementary exercises relating probability to genetics, but these will be under the enormously simplifying assumptions of no external influences and pure random mating, neither of which really hold in nature, apparently. Genetic drift is the effect of such influences on populations and their genetic make-up. It can be used at the DNA level, for example, as a kind of dating device analogous to Carbon-14: one can try to measure genetic drift chemically on the DNA chain, and use the rate of impurity or mutation to estimate when in time various significant changes took place, such as the splitting off of a new species from an old one.

4. The Poisson Process. This is very wide-ranging in its applicability. You might try to follow-up on this topic, whose surface we will barely scratch in class. This is an example, like random walk above, of a *stochastic process*, which is a probability system evolving in time.

5. Theory of Runs. This is an attempt to find out whether one can say anything about the fluctuations or runs in repeated trials of experiments. This is an elaboration of the gambler’s ruin problem which we did in class, or the experiment with coin flips we did the first week of class. You could also look at the theorem that there is no way to construct a system to beat a fair game.

6. Computer Design. There are several possibilities here, one of the most obvious is so-called queuing theory, which has to do with servers and customers, and what are the best ways for a given number of servers of a particular kind to handle a certain flow of customers efficiently. This sounds like retail marketing at first, but it enters into computer design when one considers the servers to be processors which execute commands, and the customers are the commands distributed around to processor units to be calculated or executed. One then has a lot of “customers” whom one wants to get moving along in very fast times!

7. Communications. Here one of the main concerns is the reliability of messages and what the probabilities of random errors in messages might be. There are some small examples of this scattered throughout your textbook, and some things in the last chapter which we shall not, unfortunately, have time to discuss in the course. The probabilistic idea of *entropy* arises here.

8. Failure Rate Models. These arise in engineering (random failure of a structure), or

in mortality estimations in insurance and benefit systems analysis. Many diverse models have to be used for different situations, but the most basic and important ones will be directly accessible to you (*e.g.*, the Weibull distributions, used in mechanical engineering and elsewhere).

9. *Markov Models.* A Markov chain is a simple kind of probabilistic evolution or dynamical model. The main assumption is that from step to step one always has the same transition possibilities, and these do not depend on what has happened in the past at all. An easy example of this is to flip a coin in a sequence. You will get a head or a tail in the next flip independently of anything that happened in the past. Random walk is an example of a Markov chain. Some applications give rise to “hidden Markov models”, where several possible types of Markov chains might be occurring at different times. This might occur, for example, when you are betting against the house and you do not know whether the house might not be changing the dice from roll to roll. In this case what is “hidden” is the type of dice being used, and you can observe the outcomes of the rolls, and you have to judge which Markov model is applicable when (*i.e.*, when are they using fair dice and when biased). More serious examples occur in molecular genetics, for example, where one doesn’t know where on a DNA strand the genes are located. Another example is in linguistics and speech recognition, where the “hidden process” is the type of speech someone is using, which is a behind the scenes choice which greatly influences the meaning of the sounds which are audible on the surface of things. these are both related to very interesting computer applications in terms of designs for devices which can carry out these last two tasks.

10. *Pseudo-random Number Generators.* We will see computer simulations of probability situations. *ALL* of these are based on (pseudo-)random number generators, algorithms which produce numbers subject to various conditions which are “approximately random”. Now obviously there is a contradiction in saying you have an algorithm which computes random numbers, but the idea is to produce a sequence of numbers whose rule is sufficiently complex that relations between the numbers produced will be sufficiently complex that they will not play a role in the computations to be performed. For example, if I wanted to test some arithmetic statement on some random trial numbers, and my algorithm always produced even numbers, that would not be very useful. Subtler, “higher order” relations might be acceptable.