Entropy, Martingales and Finance

18.600 Problem Set 10, due May 6

Welcome to your tenth 18.600 problem set! We'll be thinking a bit about the efficient market hypothesis, risk neutral probability, martingales, and the optional stopping theorem. These ideas are commonly applied to financial markets and prediction markets, but they come up in many other settings as well. Indeed, if X is any random variable with finite expectation, then as one observes more and more information, one's revised conditional expectation for X evolves as a martingale, and the optional stopping theorem applies to the sequence of revised expectations.

Suppose that you know you have an exactly $2^{-10} = 1/1024$ chance of dying during the next 12 months. (You can see at `https://www.ssa.gov/oact/STATS/table4c6.html` what fraction of US men and women your age die during a given year; the 1/1024 figure is way too high for women but actually a little low for men.) Now which would you prefer?

1. If you die, it happens a random time during the year (no knowledge in advance).
2. You witness the outcome of one coin toss every month over the course of ten months, and you die at the end of the year if all are heads.
3. All coins are tossed at once at the end of the year, and you die if they are all heads.

If you choose the second option, your conditional probability of dying that year will evolve as a martingale (starting at 1/1024, then jumping to 0 or 1/512, then jumping to 0 or 1/256, etc.) If you choose the third option, there is a single jump (from 1/1024 to either 1 or 0) that happens all at once. This choice posed here may seem morbid, but in fact real life poses frequent analogs of this question (involving early cancer diagnoses, mammograms, genetic tests, etc.) and the answers are not easy. How much information about our revised chances do we really want to have? How do we respond emotionally to martingale ups and downs? How do we react when the first six tosses are heads one year, and the future is suddenly scarier?

Any fan of action movies knows that the heroes frequently face circumstances where (based on non-movie logic) the conditional probability that they survive the adventure appears very low. (*C3PO: Sir, the possibility of successfully navigating an asteroid field is approximately 3,720 to 1.*) This (sequentially revised) conditional probability is like a martingale that gets very close to zero, then somehow comes back to a moderate value, then gets close to zero again, then returns to moderate, then gets *extremely close* to zero in a big climactic scene, and then somehow gets back to one. This behavior is unlikely for actual martingales. But nobody argues that the stories that get made into movies (fictional or otherwise) are typical. We like to tell the stories with close calls and happy endings, even if most stories aren't like that.

This problem set also features problems about *entropy* (an extremely important concept in (for example) statistical physics, information theory, and data compression) as well as Markov chains, which play an important role in operations research, computer science, and many other areas. If you google *seven shuffles* you'll find a famous result about how many shuffles are required to adequately mix up a deck of cards: in this case, a "shuffle" is a step in a Markov chain on the set of 52! permutations of a standard deck of cards.

A. **TEXTBOOK CHAPTER NINE:**

1. Problem/Theoretical Exercises 7: A transition matrix is said to be doubly stochastic if $\sum_{i=0}^{M} P_{ij} = 1$ for all states $j = 0, 1, \ldots, M$. Show that if such a Markov chain is ergodic, with $(\pi_0, \pi_1, \ldots, \pi_M)$ the stationary row vector, then $\pi_j = 1/(M+1), \quad j = 0, 1, \ldots, M$.

2. Problem/Theoretical Exercises 9: Suppose that whether it rains tomorrow depends on past weather conditions only through the last 2 days. Specifically, suppose that if it has rained yesterday and today, then it will rain tomorrow with probability .8; if it rained yesterday but not today, then it will rain tomorrow with probability .3; if it rained today but not yesterday, then it will rain tomorrow with probability .4; and if it has not rained either yesterday or today, then it will rain tomorrow with probability .2. Over the long term, what proportion of days does it rain?

3. Problem/Theoretical Exercise 13: Prove that if X can take on any of n possible values with respective probabilities $P_1, \ldots, P_n$, then $H(X)$ is maximized when $P_i = 1/n, i = 1, \ldots, n$. What is $H(X)$ equal to in this case?

4. Problem/Theoretical Exercise 15: A coin having probability $p = 2/3$ of coming up heads is flipped 6 times. Compute the entropy of the outcome of this experiment. (An "outcome" is a full toss sequence, e.g., $\{H, T, T, T, H, H\}$.)

5. Problem/Theoretical Exercise 17: Show that, for any discrete random variable X and function f,

$$H(f(X)) \leq H(X).$$

B. **ENTROPY EXAMPLE:** A standard die is repeatedly rolled until the first time it comes up 6. Let X be the full sequence of outcomes *up to that point*. For example, maybe $X = \{1, 4, 2, 3, 2, 5, 6\}$ or $X = \{3, 4, 2, 6\}$ or $X = \{6\}$. Compute the entropy $H(X)$. **Hint:** Let K be the length of X. Argue that $H(X) = H(X, K)$. Compute the entropy of K directly (it is a geometric random variable). Then recall the identity $H(X, K) = H(K) + H_K(X)$ from the lecture slides and find $H_K(X)$.

C. **RELATIVE ENTROPY AND WORLD VIEW:** Suppose that there are n possible outcomes of an athletic tournament. I assign probabilities $p_1, p_2, \ldots, p_n$ to these outcomes and you assign probabilities $q_1, q_2, \ldots, q_n$ to the same outcomes. If the ith outcome occurs and $p_i > q_i$ then I will interpret this as evidence that my probability estimates are better than yours, and that perhaps I am smarter than you. In short, I will feel smug. Suppose that my precise smugness level in this situation is $\log(p_i/q_i)$. Then before the event occurs, my *expected* smugness level is $\sum p_i \log(p_i/q_i)$.

1. Show that my *expected* smugness level is always non-negative, and that it is zero if and only if $p_i = q_i$ for all i. (Hint: use some calculus to find the vector $(q_1, q_2, \ldots, q_n)$ that minimizes my expected smugness level. This is rather like Problem A.1 above.)

2. Suppose that if outcome i occurs, your smugness level is $\log(q_i/p_i)$, so that your expected smugness level is $\sum q_i \log(q_i/p_i)$. We agree *a priori* that our combined smugness level will be zero no matter what (your smugness is by definition negative one times my smugness). However, you expect your smugness level to be positive (and mine to be negative) while I expect my smugness level to be positive (and yours to be negative). Try to give a short intuitive explanation for why that is.

3. Look up the term *relative entropy* and explain what it has to do expected smugness.

Remark: I expect an *infinite* amount of smugness if I assign positive probability to things that you assign zero probability. We sometimes say our probability distributions are *singular* when this is the case. As a practical matter in politics, it might be a bad thing if my professed probability distribution is close to singular with respect to yours on some of the great unknowns (e.g., likelihood that certain tax cuts help the economy or that certain public investments provide net benefits or that some religious or philosophical ideas are true). The discrepancies might make it hard for us to find *any* common political ground, even if we are both utilitarians seeking the greater good. On the other hand, discrepancies are betting/trading opportunities. If our probability differences are real (and not political smokescreen) perhaps we can make a policy bet in the form of a policy that funds your priorities if your predictions pan out and my priorities if my predictions pan out.

Remark: The *Smugness Game* is played as follows. You specify a vector q, I specify p, and when the ith outcome occurs you pay me $\log(p_i/q_i)$, which is equivalent to me paying you $\log(q_i/p_i)$. You can check that no matter what p I choose, if you know the "true probability vector" you maximize your expected payout by choosing that for q. Both players are incentivized to give their best probability assessments. Some people would like to see weather prediction and political prediction teams challenge each other to play this game.

D. **HARRIET'S MARTINGALES:** Solve the following problems.

1. Suppose Harriet has 30 dollars. Her plan is to make one dollar bets on fair coin tosses until her wealth reaches either 0 or 80, and then to go home. What is the expected amount of money that Harriet will have when she goes home? What is the probability that she will "win," i.e., that she will have 80 when she goes home?

2. Harriet uses the phrase "I think A" in a precise way. It means "the probability of the event A, conditioned on what I know now, is at least .5". Which of the following is true:

   (a) Harriet thinks she will lose.

   (b) Harriet thinks that there will be a time when she thinks she will win.

   (c) Harriet thinks that her amount of money will reach 40 and subsequently reach 20 before the game is over.

3. What is the expected number of bets she will make before reaching 0 or 80? **Hint:** Let $F(n)$ be the expected number there would be if she started with n dollars for $n \in \{0, 1, 2, \dots, 80\}$. So $F(0) = F(80) = 0$. Use a conditional expectation argument (involving value of first toss) to show $F(k) = 1 + \frac{F(k+1)+F(k-1)}{2}$, which implies $[F(k+1) - F(k)] - [F(k) - F(k-1)] = -2$. Guess a quadratic function F might be and show it is the only function with these properties.

E. **BLACK-SCHOLES:** Complete the derivation of the Black-Scholes formula for European call options, as outlined in the lecture slides, by explicitly computing

$$E[g(e^N)]e^{-rT},$$

where N is a normal random variable with mean $\mu = \log X_0 + (r - \sigma^2/2)T$ and $g(x) = \max\{0, x - K\}$.

F. **POLITICS AND ROMANCE:** David Aldous of UC Berkeley devised and told me about the following problem. How many people in a given US presidential election cycle do we expect to have their probability of becoming president *at some point* exceed 10 percent? In other words, if we look at those prediction market charts (and pretend the plots are true continuous martingales) how many candidates do we *expect* will have their number at some point exceed 10 percent? Let's consider two ways to approach this problem. (**Note:** if a *continuous* martingale like Brownian motion is below .1 at one time .1 and above .1 at a later time, then it must pass through .1 at some intermediate time. If you want to avoid thinking about continuous-time martingales, just consider a discrete-time martingale with tiny increments that has this property.)

1. Assume that at some point (well before the election) every person in the world has some small probability p to become president. The ith person has probability p_i and $\sum p_i = 1$. Assume that the ith person's probability evolves as a continuous martingale; then it has a $p_i/.1 = 10p_i$ chance to reach ten percent in the prediction markets at some time. Sum over i to get an overall expected number of people who reach this threshold.

2. Imagine a gambler who adopts the following strategy. Whenever a candidate reaches 10 percent, the gambler buys a contract on that candidate for \$10 (which will pay \$100 if the candidate wins) and holds it until the end of the election. Then at the end of the election the gambler is certain to receive \$100 (since the gambler will have purchased a contract on the winner at the first time the winner's price reached \$10). Argue that by the optional stopping theorem, the gambler makes zero money in expectation. So the expected amount of money spent on contracts must be \$100.

Here is a variant. Suppose there is a .85 chance you will get married eventually (at least once). Imagine that aliens watching your life from afar are placing bets on who your first spouse will be, and that contract prices evolve as continuous martingales. Call somebody an "almost first spouse" if at some point the market probability that this person is your first spouse exceeds .5.

3. Assuming continuity of martingales, how many almost first spouses do you expect to have over your lifetime?

4. Call a person a "serious contender" if their probability exceeds .1 at some point. How many serious contenders do you expect to have?

Remark: Does it seem a little strange that somebody who doesn't know your romantic history at all could make what appear to be substantive probabilistic assessments about your future love life? Also, did I need to bring aliens into this story? Could it be your friends or parents, or maybe you yourself, assessing these probabilities?

Remark: I have no data on this, but I have heard it argued that people in relationships are irrational, because their subjective relationship-viability estimates fluctuate more than martingale theory predicts. If your estimate of the probability you will marry your current significant other regularly alternates between over .8 and below .2, and you expect this to happen several more times, then your evolving subjective probability estimates will not (according your current subjective probability) evolve as a martingale. This puts you in violation of economic rationality assumptions. Assuming rationality/consistency, the probability measure *you* assign to *your own* future revised-probability trajectory should make it a martingale. A one-step example of a violation: "Today I think there is an 80 percent chance we'll marry some time in 2018, but my *expectation* of what my probability will be after tomorrow's date is 90 percent." Of course, none of us is *actually* fully rational in the sense of having consistent and well defined probabilities for all future outcomes.

G. **RISK NEUTRAL PROBABILITY DEPENDS ON CURRENCY:** Imagine that at this particular moment on the currency market, one dollar has the same value as one euro. Let R be the event that (at some time during the next year) the euro rises in value relative to the dollar, so that the euro becomes worth two dollars. Let P_d be the cost, in dollars, of a contract that gives you one dollar if and when the event R occurs. Let P_e be the cost, in euros, of a contract that gives you one euro if and when event R occurs. Argue that one should expect $P_e = 2P_d$.

Remark: Assuming no interest, you can interpret P_d as the risk neutral probability of R (using dollars as the *numéraire*), but you can also interpret P_e as the risk neutral probability of R (using euros as the *numéraire*). This should convince you that the risk neutral probability of R cannot *always* be interpreted as a "general consensus of the subjective probability that R will occur". (That interpretation is only reasonably if the value of money does not depend on whether R occurs.) In October of 2016, currency traders believed (correctly, it turns out) that the price of the Mexican peso versus the US dollar would fall significantly if Trump were elected. Based on this, the risk neutral probability of Trump's election should have been *lower* with pesos as the numéraire than with dollars as the numéraire. (Imagine the extreme case: if it were known that Trump's election would make pesos worthless, then the price — in pesos — of a contract paying a peso if Trump were elected would be zero.)

Concluding remark: Congratulations on finishing (or at least reading to the end of) your final problem set! Your problem sets and the remarks therein have *briefly* introduced you to many topics: Powerball odds, Occam's razor, hypothesis testing, the Doomsday argument, p-values, Siegel's paradox, subprime lending, modern portfolio theory, diversification, the capital asset pricing model, idiosyncratic versus systemic risk, utility and risk aversion, Poisson bus inefficiency, Gompertz mortality, radioactive decay and half life, Cohen's d, clinical trials, open primary voting, Pascal's wager, infinite expectation paradoxes, correlation versus causation, the Kelly strategy, least squares regression, regression to the mean, publication bias, relative entropy, the Black-Scholes derivation, and the dependence of risk neutral probability on the numéraire.

Take a moment to review the problem sets and recall the stories you have forgotten. Topics that appear only in problem sets (not in lecture notes or practice exams) are not likely to be on your final. I nonetheless hope you review and retain at least some understanding of these stories. You could not have understood them without the math you have learned, and reviewing them may help you solidify your math skills and integrate probability into your thinking about the world. I also hope some of you use probability to solve big problems: to fundamentally improve our approach to science, medicine, criminal justice, economics, engineering, and so forth. Failing that, I hope you'll at least have fun with all of this.

Best of luck!

Okay, one more (optional) paradoxical remark: The following is known in philosophy circles as the *Sleeping Beauty problem*. On Sunday night a woman agrees to the following experiment. She will be put into a medically induced sleep. Then a fair coin will be tossed. If the toss comes up heads, she will be woken and interviewed *both Monday and Tuesday* mornings before being returned to sleep (each time with an amnesia-inducing drug that makes her forget the experience). If the coin comes up tails, she will be woken and interviewed *only on Monday* morning. Either way, she awakes without interview on Wednesday and the experiment is over.

When she is interviewed (on either Monday or Tuesday), she does not know either what day it is or what the coin outcome was. So we can think of $\{Mon, Tue\} \times \{H, T\}$ as the sample space corresponding to her uncertainty at that point. Basic symmetries suggest that during any interview she interprets the states $\{Mon, H\}$ and $\{Tue, H\}$ as being equally likely, and also that she interprets and $\{Mon, H\}$ and $\{Mon, T\}$ as being equally likely; hence she sees each as having probability $1/3$. (She knows that $\{Tue, T\}$ isn't possible if she is awake.)

But here's the paradoxical part. On Sunday night, she thinks there is a $1/2$ chance that the coin is heads, but she also *knows* that on Monday morning, according to her updated probability estimate, she will think there is a $2/3$ chance. At first glance this seems to violate the fact that sequentially revised conditional expectations evolve as martingales. How can she rationally think there is a $1/2$ chance now but also *know* that she will think there is a $2/3$ chance tomorrow morning?

Mathematically, the problem is that we cannot say she is just revising her probability by conditioning on new information. She is also *forgetting* and/or *losing track of time* in a way that creates a new kind of uncertainty. The day of the week is not an unknown from the woman's point of view on Sunday, but it *becomes* unknown as of Monday morning. There have been over a hundred philosophy written papers about this problem, some of which reach different conclusions about the Monday morning heads probability or "credence." See `https://www.quantamagazine.org/solution-sleeping-beautys-dilemma-20160129/` or look up the survey by the mathematician Peter Winkler. (You can also google "sleeping beauty problem" and read the Wikipedia artice.) For a simpler "forgetfulness" story, imagine you know that a coin toss just came up heads, but you also know that a year from now you'll have forgotten this fact, so that at that point you'll subjectively think the chance the coin was heads is .5. Your sequentially revised heads probability cannot be a martingale in this story, since it violates the optional stopping theorem. Forgetting is problematic.

So... try not to forget too much of what you learned in this course. And best of luck again!