Expectation, covariance, binomial, Poisson

18.600 Problem Set 4, due October 4

Welcome to your fourth 18.600 problem set! The interesting topics we have discussed in lecture include the linearity of expectation, the bilinearity of covariance, and the notion of *utility* as used in economics. (Under certain "rationality" assumptions everyone has a utility function whose expectation they seek to maximize.) We will see in this problem set how these ideas play a role in some important (though perhaps overly simplistic) theories from finance (MPT and CAPM). We will have more on binomial/Poisson random variables and a chance to learn about Siegel's paradox.

Please stop by my weekly office hours (2-249, Wednesday 3 to 5) for discussion.

A. FROM TEXTBOOK CHAPTER FOUR:

1. Problem 23: You have \$1000, and a certain commodity presently sells \$2 per ounce. Suppose that after one week the commodity will sell for either \$1 or \$4 an ounce, with these two possibilities being equally likely.

   (a) If your objective is to maximize the expected amount of money that you possess at the end of the week, what strategy should you employ?

   (b) If your objective is to maximize the expected amount of the commodity that you possess at the end of the week, what strategy should you employ?

Remark: Look up Siegel's paradox. It's pretty interesting.
`http://mindyourdecisions.com/blog/2012/03/15/siegels-paradox-about-exchange-rates/`

B. On ACT Planet, Jack is preparing to take a test called the Math ACT. Jack knows his stuff but is error prone under pressure, and because of this he only gets the right answer 85 percent of the time. His success probability is the same for all problems (no matter how hard they are for others) and his outcomes are independent from one problem to another. If he gets his expected 51 out of 60 answers correct, it comes to a Math ACT score of 30. (On ACT planet, the score conversion table is the same each time the test is given.)

Jack wants to attend "Thirty Four or Higher University" (known as TFOHU) for which he needs a 34 on the Math ACT, which requires at least 55 out of 60 correct. Fortunately, the university only requires that he obtain that score once. He is not required to report scores that fall below the treshold. Jack decides to invest the time and money to take the exam 12 times. Assuming Jack's abilities remain constant, what is the probability that he gets a sufficiently high score (i.e., at least 55 correct answers) at least once? Give both an algebraic expression and an approximate percentage. (A calculator like `http://stattrek.com/online-calculator/binomial.aspx` may help you compute the numerical value.)

C. Suppose that during each minute there is a 1 in $500,000$ probability that there is an accident at a particular intersection (independently of all other minutes). Using the approximation of $500,000$

minutes per year, we expect to see 1 accident per year on average. One year somebody proposes to install a new kind of stoplight to reduce accidents. You believe *a priori* that there is a 1/3 chance that the new stoplight is *effective*, in which case it will reduce the accident rate by fifty percent, and a 2/3 chance it will have no effect. The new stoplight is installed and during the next year there are no accidents. Using Poisson approximations, compute your updated estimate of the probability that the light is effective.

Remark: It is often hard to tell whether preventative measures against rare events are having an effect. With twenty years of data we might be more confident, but by that point accident rates may have changed for other reasons (e.g., self driving cars). On the other hand the $k!$ in the Poisson denominator means that *large* numbers are *extremely* unlikely. If we suddenly see 10 accidents in one year, we should seriously question our assumption that the number is Poisson with $\lambda = 1$ or $\lambda = 1/2$.

D. Larry the Very Subprime Lender gives loans of size \$10,000. In 25 percent of cases, the borrower pays back the loan quickly with no interest or fees. In 50 percent of cases, the borrower disappears (moves away, declares bankruptcy, dies) without paying anything. In 25 percent of cases, the borrower pays back the loan slowly and — after years of ballooning interest payments, hefty fees, etc. — pays Larry a total of \$100,000. However, in this scenario, Larry has to give \$60,000 to third parties (repo services, foreclosure lawyers, eviction teams, bill collectors, etc.) in order to get the borrower to pay the \$100,000. Compute the following:

(a) The expectation and variance of the *net* amount of profit Larry makes from each loan (after subtracting collection expenses and the initial \$10,000 outlay).

(b) The expectation and variance of the *net* amount a given borrower ends up paying (i.e., amount paid minus amount borrowed).

Note: You might have ethical concerns with Larry's (unrealistic) business model. Google *payday loan* or *buy here pay here* or (for purely negative side) *predatory lending* (maybe start with the Wikpedia articles) if you want more realistic information about the legal and moral issues involved in lending to populations with high default risk. It is a large and complicated subject.

E. Define the covariance $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$. Define the *corelation* between X and Y to be $\mathrm{Cov}(X, Y)/\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}$. One can show that the correlation between any two random variables is always between -1 and 1. *Very* roughly speaking, the correlation is high (i.e., close to 1) if X tends to be high when Y is high and tends to be low when Y is low.

1. Check that $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$, that $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$, and that $\mathrm{Cov}(\cdot, \cdot)$ is a bilinear function of its arguments. That is, if one fixes one argument then it is a linear function of the other. For example, if we fix the second argument then for real constants a and b we have $\mathrm{Cov}(aX + bY, Z) = a\mathrm{Cov}(X, Z) + b\mathrm{Cov}(Y, Z)$.

2. If $\mathrm{Cov}(X_i, X_j) = ij$, find $\mathrm{Cov}(X_1 - X_2, X_3 - 2X_4)$.

3. If $\mathrm{Cov}(X_i, X_j) = ij$, find $\mathrm{Var}(X_1 + 2X_2 + 3X_3)$.

4. Suppose that V and $X_1, X_2, \dots, X_n$ are random variables, that $\mathrm{Var}(V) = 1$, that $\mathrm{Cov}(X_i, V) = b_i$ for each i and that $\mathrm{Cov}(X_i, X_j) = c_{i,j}$ for each pair i, j, where b_i (for $1 \le i \le n$) and $c_{i,j}$ (for $1 \le i, j \le n$) are known constants. Suppose for some fixed constants $a_1, a_2, \dots, a_n$, we write $X = \sum a_i X_i$. Then demonstrate the following:

   (a) $\mathrm{Cov}(X, V) = \sum_{i=1}^n a_i b_i$

   (b) $\mathrm{Var}(X) = \sum_{i=1}^n \sum_{j=1}^n a_i c_{i,j} a_j$.

   (c) The correlation between X and V is $\frac{\sum a_i b_i}{\sqrt{\sum a_i c_{i,j} a_j}}$.

   With some calculus and linear algebra (which I won't make you do) you can use the above to find a choice of $a_1, a_2, \dots, a_n$ that *maximizes* the correlation between X and V.

Remark: Many university rating systems (and clickbaity lists like "top ten cities for singles" or "best companies to work for") are constructed using a weighted sum X of measurements X_i each believed to be *correlated* with some (hard to define) *overall* value V. For example, US News measures what fraction of alumni donate, how much professors are paid, what fraction of faculty have PhDs, what fraction of students were top ten percent in high school, etc. In each case, the measured quantity is not something students necessarily care about *for its own sake* — rather, it is believed to be *correlated* with things they care about. The QS World University Rankings (where MIT is first) use a weighted sum of six different quantities (citations per faculty, academic and employer reputation, etc.)

Many feel that these rankings are useful in holding universities accountable and conveying at least a rough sense of where the strong universities are. Others are more critical. One problem is that the rankings depend heavily on the choice of measured quantities (the X_i) and the weights (the a_i). These choices often seem arbitrary and *ad hoc*, and differ greatly from one ranking system to another. Another problem is that even if we pretend there is a quantity V that represents *overall value*, and even if we have defined an X such that the correlation between X and V is high (say .8) across all universities, it is not clear that the correlation remains high if we restrict attention to, say, the top 20 universities. (Maybe a statistic like "5 $*$ *height in inches* minus 2 $*$ *age in years*" is well correlated with basketball ability in a randomly chosen adult, but not in a randomly chosen NBA player.) A final concern is that institutions may "game the system" by taking actions that increase X without increasing V. Some worry that these actions waste resources, and may even decrease V. (The adage that "When a measure becomes a target, it ceases to be a good measure" is sometimes called Goodhart's law.) Google *us news rankings controversy* for some seriously anti-ranking polemics.

F. Use the following identities (some more well known than others) to solve the problems below:

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} \qquad \sum_{n=0}^{\infty} \frac{1}{n!} = e \qquad \sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n} = \log(2) \qquad \sum_{n=2}^{\infty} \binom{n}{2}^{-1} = 2$$

$$\sum_{n=1}^{\infty} n^s = \prod_{p \text{ prime}} (1-p^s)^{-1} \text{ if } s < -1 \qquad \frac{1}{\pi} = \frac{\sqrt{8}}{99^2} \sum_{k=0}^{\infty} \frac{(4k)!(1103 + 26390k)}{(k!)^4 396^{4k}}$$

1. A town launches an annual marathon. Let M_n and W_n denote the winning times for men and women during the nth year. Assume the M_n are independently chosen from some continuous probability measure and the W_n are independently chosen from another continuous probability measure (the precise choice doesn't matter). Observe that the probability that there is a women's record during the nth year (i.e., that $W_n < W_k$ for all $k < n$) is $1/n$.

   (a) Compute the expected number of years during which there is simultaneously a men's record and a women's record (assuming the annual marathon will continue forever).

   (b) Compute the probability that there is *never* simultaneously a men's record and women's record during a *prime* numbered year.

   (c) Let K be the number of the first year during which there is *not* a women's record. (So $K \geq 2$.) Compute the expectation of K. You can use the fact that $E[K] = \sum_{k=0}^{\infty} P[K > k]$.

   (d) Let L be number of year when the first woman winner first sees her record beaten. Compute the probability L is even. Hint: observe that $P(L \text{ even}) = P(L > 1) - P(L > 2) + P(L > 3) - P(L > 4) + \ldots$.

   (e) Compute the expected number of years $n \geq 2$ when the most recent two winning men are fastest, i.e. $\max\{M_n, M_{n-1}\} < M_k$ for $k < n - 1$.

2. In a future unified muggle/wizard sorting process, an infinite line of people is waiting to be sorted. At each step, the sorting hat independently declares a person to be a wizard with probability 1/99, in which case it sorts the individual into one of four houses (each with probability 1/4). In between sortings, if it is ever determined that

   (a) everyone sorted thus far has been a wizard, and

   (b) each of the four wizard houses has an equal number of people,

   then this rare occurrence is celebrated in the obvious way: namely, the sorting is paused, an interhouse Quidditch tournament is conducted, and each person in the winning house receives 26390 Galleons, while the house itself receives a trophy worth 1103 Galleons. (A trophy is also awarded during the "empty" game that occurs at the beginning when all houses have zero people.) Compute the expectation of the total value that will ever be awarded (trophies included).

Remark: These problems are from a garageband clip about identities I posted last year `http://math.mit.edu/~sheffield/2018600/kindofthing.mp4`. (Most mathematical music videos don't contain much math; this is an exception, for better or worse.) You can use the answers at the end of the clip to check your numbers, but you need to show work to get credit (i.e., at least write in a few words what the individual terms/factors represent). If you think events like $E_p = \{\text{no record occurs in } p\text{th year}\}$ are *independent* you should say why. **Hint:** Fix n and let σ be the permutation such that year j was $\sigma(j)$th fastest year for the women's time (among the first n years). Start by arguing by symmetry that all such permutations are equally likely.

G. Instead of maximizing her expected wealth $E[W]$, Jill maximizes $E[U(W)]$ where $U(x) = -(x - x_0)^2$ and x_0 is a large positive number. That is, Jill has a *quadratic utility function.* (It may seem odd that Jill's utility declines with wealth once wealth exceeds x_0. Let us assume x_0 is large enough so that this is unlikely.) Jill currently has W_0 dollars. You propose to sample a random variable X (with mean μ and variance σ^2) and to give her X dollars (she will lose money if X is negative) so that her new wealth becomes $W = W_0 + X$.

1. Show that $E[U(W)]$ depends on μ and σ^2 (but not on any other information about the probability distribution of X) and compute $E[U(W)]$ as a function of x_0, W_0, μ, σ^2.

2. Show that given μ, Jill would prefer for σ^2 to be as small as possible. (One sometimes refers to σ as *risk* and says that Jill is *risk averse.*)

3. Suppose that $X = \sum_{i=1}^n a_i X_i$ where a_i are fixed constants and the X_i are random variables with $E[X_i] = \mu_i$ and $\mathrm{Cov}[X_i, X_j] = \sigma_{ij}$. Show that in this case $E[U(W)]$ depends only on the μ_i and the σ_{ij} (but not on any other information about the joint probability distributions of the X_i) and compute $E[U(W)]$. Hint: first compute the mean and variance of X.

Remark: We conclude (assuming quadratic utility) that portfolio builders care *only* about expectations and covariances of items in their portfolio. This idea underlies the (1990 Nobel Prize Winning) *Modern Portfolio Theory* (MPT) and *Capital Asset Pricing Model* (CAPM). Before these theories, it was believed that when the *variance* of an asset return is high, the *expected* return should be higher as well (the *risk premium*) because otherwise people wouldn't buy risky assets. MPT and CAPM predict that one gets a risk premium for *systemic risk* (the part of the variance explained by correlation with the *market portfolio*, defined to be the sum total of all risky assets) but not for *idiosyncratic risk* (exposure to which can be reduced by diversification). These theories also predict that everyone's optimal investment strategy is to put some (investor-dependent) fraction of their money in a risk-free asset and the remainder in the market portfolio (which we think of as a giant index fund). Google MPT and CAPM to read about how well or poorly these theories match reality.

4. Suppose that $X_1, X_2, \ldots, X_n$ are independent random variables with the same mean and variance. Show that among all random variables of the form $\sum_{i=1}^n a_i X_i$ (where the a_i are non-negative numbers with $\sum_{i=1}^n a_i = W$ for some fixed constant W) the one with the smallest variance is the one with $a_i = W/n$ for each i.

Remark: If the (presumed i.i.d.) X_i are returns on n investments, then the above implies that one minimizes risk by dividing wealth equally among the investments. In the story below, the X_i are the overall stock market returns in different *years*. Suppose you plan to contribute K dollars to your child's college fund annually for 18 years, dividing wealth between a (zero interest) safe investment and a (risky) stock index fund. You decide in advance that on the ith year, you will invest a_i dollars in stocks. Then (if $\sum a_i$ is held constant) your variance is minimized if you invest the same amount in stocks each year. If you instead keep a fixed *percentage* of wealth in stocks each year, then your final value will depend most heavily on market performance during the later years (when you have the most money in the account). This is why some financial planners recommend being more *aggressive* (i.e., open to higher risk in exchange for higher expectation) during early years and more *conservative* as you get closer to using the money. *However,* if the money were all contributed *upfront* with no annual K-dollar influx (and you assumed *logarithmic* utility) you could make a case for keeping a fixed *percentage* in stocks. More on this later.

Remark: People often say utility functions should be strictly concave (negative second derivative) to explain risk aversion... but is that necessarily true? Here is a naive story about charitable giving. Suppose your utility function is given by your own health/comfort plus a constant c times the sum of the health/comfort of all other humans on the planet. For example, if $c = .01$, then you are mostly selfish, but you would be willing to give up a comfort for yourself if it would enable more than 100 strangers to enjoy the same comfort. You'd give up your life if you could save more than 100 other lives. Utilitarians might theorize that it is a good to thing that $c > 0$ (so that we help others when we can make a big difference) but maybe also a good thing that $c < 1$ (since a little selfishness might be efficient in practice). As you acquire more money, there may come some point at which you believe that the marginal value of another dollar to you (in added health/comfort) is *less than* c times the amount a dollar donated to a global charity with relatively high expected impact (like those profiled at givewell.org) would increase health/comfort for others. After that point, in principle you should donate *all* of your additional money to charity. If this is indeed your plan, then your utility function might be very close to linear for a long time after that point, since the amount of good you do in a huge global effort is roughly linear in the amount you give.

Remark: Some economists say that in reality charitable giving should be modeled as a consumptive good (that happens to have positive externality — google *"warm glow giving"*) that has to compete with other consumptive goods among even the very wealthy. This point of view might predict actual behavior better than what I sketched above.

Remark: A "rational" person (in the economic sense) has a utility function and a subjective probability measure, and makes decisions that optimize expected utility. But Arrow's impossibility theorem (look it up) states that (under any reasonable voting scheme) a democratic *group* may prefer A to B and B to C and C to A, which would imply $U(A) > U(B) > U(C) > U(A)$ (contradiction) if the *group* had a utility function. Political parties, companies, and entire countries can all be "irrational" to a greater extent than their individual members.