

Cauchy, beta, gamma and the infinite expectation paradox

18.600 Problem Set 7, due April 8

Welcome to your seventh 18.600 problem set! This problem set will feature problems about beta, Γ , and Cauchy random variables. These random variables are not quite as ubiquitous as others we have discussed (exponential, uniform, normal, Poisson, binomial) but they are fun and they do come up. The problems should help you internalize both the definitions and some standard interpretations.

But first let's warm up with some stories. Imagine that you live in a country with hundreds of millions of people infected by a deadly plague. An eccentric (but completely credible) doctor makes an announcement. In one gymnasium, the doctor says, 1,000,000 independent fair coins will be tossed as randomly as possible; let H_1 be the number that come up heads. In a second gymnasium 1,000,002 coins will be similarly tossed; let H_2 be the number that come up heads. The society must select in advance a number $j \in \{1, 2, \dots\}$. After the tosses, the doctor pledges that $100H_j$ plague sufferers will be given treatment and cured. The interesting thing about this (admittedly absurd) scenario is the following. On one hand, the probability that $H_1 > H_2$ is extremely close to $1/2$ — so you cannot predict which choice will look better *in retrospect*. On the other hand, since you understand the probabilities perfectly, you *know* that $j = 2$ is the *better bet*, and that choosing $j = 2$ over $j = 1$ will save, in expectation, 100 human lives.

Related story: suppose that on your morning commute you face the difficult decision of whether to punch a fellow commuter in the face. If you imagine two parallel universes — one in which you throw the punch, and one in which you don't — they would probably start to look very different after years of cascading effects. One universe might have wars that the other did not, etc., and it is hard to predict which would be better in the long term. On the other hand, one universe involves a predictable set of bruises and battery charges that the other does not, so *based on what you know now* you assign the “don't punch” option (like the $j = 2$ option) a higher expected utility.

These stories illustrate that even among rational utilitarians, one's degree of conviction that it is better to choose A than B does not exactly correlate with one's assessment of the probability that A will yield a better long term outcome than B. Compare the question “How convinced are you that, given what we know, burning less fossil fuel now is a *better bet* for humanity?” to the question “How certain are you that burning less fossil fuel now would produce a better future than the alternative?” I won't address either question except to remark that in principle the answer could be “Very” to the first and “Not at all” to the second. You can ask similar questions when deciding what insurance to buy, who to vote for, and so forth. In each case, you might proclaim confidence in your decision while confessing ignorance as to how it will play out.

Final musing: many of you are familiar with *Pascal's wager*. The general idea is that if choosing A over B comes with a finite cost but a positive probability (however small)

of an infinite payoff, then one should always choose A. Pascal's conclusion was that if living a virtuous life leads (with even a tiny probability) to an eternal reward, then it is a worthwhile sacrifice to make. A common criticism is that this kind of thinking can lead to violence (killing heretics who *might* lead souls stray, or dissidents who *might* obstruct an endless Marxist utopia) as well as virtue. A more mathematical concern is that in principle there may be many choices, each of which we expect to do an infinite amount of good (and perhaps also an infinite amount of harm) and that there is no obvious mathematical way to compare the competing infinities.

The comparison difficulties associated with infinite expectations can arise even when the payoffs themselves are finite with probability one (e.g., if the utility payout is a Cauchy random variable). This problem set illustrates this point with a particularly vexing form of a famous envelope switching paradox. Interestingly, in this paradox, the conditional expectations used for decision making are all finite; but a certain *a priori* expectation is infinite, and that is the root of the paradox. I hope that you enjoy thinking about the story, and that it causes you at most a finite amount of existential angst.

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

A. FROM TEXTBOOK CHAPTER FIVE:

1. Theoretical Exercise 26: If X is a beta random variable with parameters a and b show that

$$E[X] = \frac{a}{a+b},$$

$$\text{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}.$$

2. Theoretical Exercise 28: Consider the beta distribution with parameters (a, b) . Show that
 - (a) when $a > 1$ and $b > 1$, the density is unimodal (that is, it has a unique mode) with mode equal to $(a-1)/(a+b-2)$;
 - (b) when $a \leq 1, b \leq 1$, and $a+b < 2$, the density is either unimodal with mode at 0 or 1 or U -shaped with modes at both 0 and 1;
 - (c) when $a = 1 = b$, all points in $[0, 1]$ are modes.

B. FROM TEXTBOOK CHAPTER SIX:

1. Problem 30: Jill's bowling scores are approximately normally distributed with mean 170 and standard deviation 20, while Jack's scores are approximately normally distributed with mean 160 and standard deviation 15. If Jack and Jill each bowl one game, then assuming that their scores are independent random variables, approximate the probability that

- (a) Jack's score is higher;
 - (b) the total of their scores is above 350.
2. Theoretical Exercise 12: Show that the jointly continuous (discrete) random variables X_1, \dots, X_n are independent if and only if their joint probability density (mass) function $f(x_1, \dots, x_n)$ can be written as

$$f(x_1, \dots, x_n) = \prod_{i=1}^n g_i(x_i),$$

for nonnegative functions $g_i(x)$, $i = 1, \dots, n$.

C. FROM TEXTBOOK CHAPTER SEVEN:

1. Theoretical Exercise 19: Show that if X and Y are identically distributed (but not necessarily independent) then $\text{Cov}(X + Y, X - Y) = 0$.

D. Let p be the fraction of MIT students who love Taylor Swift — or, more precisely, the fraction who will *say* they love Taylor Swift when you ask (making it clear that you *absolutely* require a simple yes or no answer). Let's make believe that your initial Bayesian prior for p is uniform on $[0, 1]$. Now ask three of your fellow students (actually do this!) one at a time whether they love Taylor Swift, and write the pair (# yes answers so far, # no answers so far) before you start and after each time you ask a question. For example, you will write the pairs

$$(0, 0), (1, 0), (2, 0), (3, 0)$$

if everyone you ask loves Taylor Swift. Pretend that you have chosen your people uniformly at random from the large MIT population, so that each answer is yes with probability p and no with probability $(1 - p)$ independently of the other answers. Then write down each of the four number pairs, and beside each one draw a rough picture of the graph of the revised probability density function for p that you would have at that point in time, along with its algebraic expression, which should involve powers of p and $(1 - p)$. You can use graphing software if you want. Beside each graph write down the corresponding conditional expectation for p (using the results from part A) given what you know at that time.

E. The following is one formulation of a famous “two envelope” paradox. Jill is a money-loving individual who, given two options, invariably chooses the one that gives her the most money in expectation. One day Harry, a trusted (and capable of delivering) individual, offers her the following deal as a gift. He will secretly toss a fair coin until the first time that it comes up tails. If there are n heads before the first tails,

he will place 10^n dollars in one envelope and 10^{n+1} dollars in the second envelope. (Thus, the probability that one envelope has 10^n dollars and the other has 10^{n+1} dollars is 2^{-n-1} for $n \geq 0$.) Harry will then hand Jill the pair of envelopes (randomly ordered, indistinguishable from the outside) and invite her to choose one. After Jill chooses an envelope she will be allowed to open it. Once she does, she will be allowed to either keep the money in the first envelope or switch to the second envelope and keep whatever amount of money is in the second envelope. However, if she decides to switch envelopes, she has to pay a one dollar “switching fee.”

- (a) If Jill finds 100 dollars in the first envelope she opens, what is the conditional probability that the other envelope contains 1000 dollars? What is the conditional probability that the other envelope contains 10 dollars?
- (b) If Jill finds 100 dollars in the first envelope she opens, how much money does Jill expect to win from the game if she does not switch envelopes? (Answer: 100 dollars.) How much does she expect to win (net, after the switching fee) if she *does* switch envelopes?
- (c) Generalize the answers above to the case that the first envelope contains 10^n dollars (for $n \geq 0$) instead of 100.
- (d) Jill concludes from the above that, no matter what she finds in the first envelope, she will expect to earn more money if she switches envelopes and pays the one dollar switching fee. This strikes Jill as a bit odd. If she knows she will always switch envelopes, why doesn't she just take the second envelope first and avoid the envelope switching fee? How can she be maximizing her expected wealth if she spends an unnecessary “switching fee” dollar no matter what? How does one resolve this apparent paradox?

F. Alice and Bob are interested in having a child and, after difficulty conceiving, decide to undergo a medical procedure called IVF. In their universe, each couple has a random quantity p , uniformly distributed on $[0, 1]$, which indicates the probability that they will conceive a child after a cycle of IVF treatment. (The value p depends on permanent biological characteristics of Alice and Bob, but its value is unknown to them, so we model it as a random variable.) If Alice and Bob attempt multiple cycles, each one succeeds with the same probability p , independently of what happens on previous cycles.

- (a) Explain intuitively why (in this universe) the probability that Alice and Bob conceive after one cycle should be .5 (i.e., the expected value of p).
- (b) *Given* that Alice and Bob did not conceive during the first $(k - 1)$ cycles, what is the updated Bayesian probability density for the random variable p ?

- (c) Use the answer in (b) to explicitly compute the expected value for p , given that the couple did not conceive during the first $(k - 1)$ cycles. The answer is the conditional probability that the couple conceives during the k th cycle, given that they did not conceive during the first $(k - 1)$ cycles. (One can prove in general that if one *first* chooses r in some random fashion and *then* tosses a coin that is heads with probability r , the overall probability of heads is the expectation $E[r]$.)
- (d) Compute the conditional probability describe in (c) in a different way: imagine that $X_0, X_1, X_2, \dots, X_k$ are uniformly and independently distributed on $[0, 1]$. Write $p = X_0$ and declare that the j th cycle succeeds if and only if $X_j < X_0$. Show that this model is equivalent to the one initially described, and then explain why the probability that $X_k < X_0$, *given* that X_0 is the smallest of the set $\{X_0, X_1, \dots, X_{k-1}\}$, should be $1/(k + 1)$. [Hint: use symmetry to argue that *a priori* the rank ordering of X_0, X_1, \dots, X_k is equally likely to be given by each of the $(k + 1)!$ possible permutations.]
- (e) Suppose that instead of being uniform the random variable p is *a priori* distributed on $[0, 1]$ according to the density function $f(x) = 2 - 2x$. (This might be more realistic, see remark below.) Under this assumption, compute the probability of success on the k th cycle given that the first $(k - 1)$ cycles failed. [Hint: recognize $f(x)$ as itself a beta random variable and reduce to the previous case.]

Remark: This problem was inspired by a NY Times article called *With in vitro fertilization persistence pays off* (look it up) which reports on a large study:

The rate of live births for participants after the first cycle in the new study was 29.5 percent, compared with 20.5 percent after the fourth cycle, 17.4 percent after the sixth cycle, and 15.7 percent after the ninth cycle.

The numbers start a bit below our answer in (e) (since $.295 < 1/3$) and end up larger (since $.157 > 1/11$). This may suggest that p values are not distributed according the f that we guessed (somewhat arbitrarily) in (e). On the other hand, maybe different people have different Bayesian priors for p (based on age, known physical issues, etc.) and those whose p values are expected *a priori* to be small tend to discontinue IVF after fewer cycles; if so, this could explain the higher reported success rates for later cycles.

G. Suppose X_1, X_2, \dots, X_6 are independent Cauchy random variables. Compute the probability that $X_1 + X_2 + X_3 > X_4 + X_5 + X_6 + 3$. (Hint: try combining the spinning flashlight story with left-right symmetry and the fact that the average of independent Cauchy random variables is itself a Cauchy random variable.)

H. Let X be a Γ random variable with parameters $\lambda = 1$ and $\alpha = n$ where n is an integer. Let Y be an exponential random variable with parameter $\lambda = 1$. Derive the variance for X from the variance for Y using a “waiting time until n th bus” story.