Statistics 531: Probability Theory
A Semester Long Problem List

INSTRUCTIONS Everyone should spend a subtantial amound of time time with the problems. Some are easy, but many are not. You can feel free to work with others on these problems.

If you find a solution you like, you or your team can write it up as a *Toolbox Essay* and share your solution with the class. I will not be collecting the solutions to these problems, so you can work on them at your own pace. You will want to take care to manage your time; it is easy to fall behind unless you have good, regular work habits.

At least a few of these problems can be expected to have clones that will be on the mid-term or the final. More details about these problems will evolve over the semester.

Send me mail if you suspect a bug, or if you want suggestions about writing a *Toolbox Essay.*

NOTATION: $\mathbb{I}(A)$ is the the indicator of the event A.

PROBLEM 1.BAD TO WORSE. Suppose $(\Omega, \mathcal{F}, P)$ is a Borel probability space, or, equivalently, suppose the probability space is $([0,1], \mathcal{B}, \lambda)$ where $\mathcal{B}$ is the set of Borel subsets of $[0,1]$ and λ is Lebesgue measure. Suppose the random variables X_n are non-negative and $E[X_n] \to \infty$ as $n \to \infty$. Can you find events A_n, $n = 1, 2, ...$ such that

(1) $P(A_n) \to 0$ as $n \to \infty$, and
(2) $E[X_n \mathbb{I}(A_n)] \to \infty$ as $n \to \infty$.

PROBLEM 2. GET A MAX FOR FREE. Suppose that X_i, $i = 1, 2, \ldots$ are i.i.d. and have mean $\mu > 0$. Let $S_k = X_1 + X_2 + \cdots + X_k$. Show that we have with probability one that

$$\lim_{n\to\infty} \frac{1}{n} \max_{1\le k\le n} S_k = \mu.$$

PROBLEM 3. UNIFORM INTEGRABILITY AND A SPECIAL SUBSEQUENCE. Assume that the sequence $\{X_n : n = 1, 2 \ldots\}$ is uniformly integrable. Show that there is a subsequence $\{X_{n_k} : i = 1, 2 \ldots\}$ with the following property:

For every bounded random variable Y, the sequence $E[Y X_{n_k}]$ of real values converges as $k \to \infty$.

PROBLEM 4. TWO FOR ONE. Consider a probability space $(\Omega, \mathcal{F}, P)$ and assume that the collection C of random variables is uniformly integrable. Now consider the class

$$C^* = \{Y : Y = E(X \,|\, \mathcal{G}) \text{ for some } X \in C \text{ and some sub-sigma field } \mathcal{G} \subset \mathcal{F}\}.$$

Show that C^* is uniformly integrable. (Note: In class we did the version where C is just a single random variable.)

As a warm-up for the next problem, you may want to remind yourself of *Lusin's Theorem*, which says that if $(\Omega, \mathcal{F}, P)$ is a probability space and X_n converges almost surely to X, then for all $\epsilon > 0$, there is a measurable set Ω_ϵ such that $P(\Omega_\epsilon) \geq 1 - \epsilon$ and X_n converges uniformly to X for all $\omega \in \Omega_\epsilon$.

PROBLEM 5. DIAGONAL CONVERGENCE. Now suppose $\{X_n\}$ is a uniformly integrable sequence and assume that $\mathcal{F}_n$ is an increasing sequence of σ-fields. Suppose that X_n converges almost surely to X.

(a) Show that as $n \to \infty$, we have "diagonal convergence":

$$\lim_{n\to\infty} E[X_n \,|\, \mathcal{F}_n\,] = E[X \,|\, \sigma\{\cup_n \mathcal{F}_n\}] \quad \text{in } L^1, \tag{1}$$

(b) Show by an example that one the L^1 convergence in (1) cannot be replaced with "convergence with probability one."

Note 1. In the case when $X_n = X$ for all n, the limit (1) is part of Levy's theorem. As a consequence, this result may suggests a "generic" way to generalise certain theorems; one passes from the "fixed" case to the "diagonal case". You might want to use this idea to invent a new result that you can write up as a Toolbox essay.
Note 2. You can find this problem in some books, but the proof you find may be sketchy, incomplete, or otherwise not-too-cool. Do a careful job with this one, including the identification of the limit with the help of the π-λ theorem, or some similar tool.
Note 3. It has taken me three passes to get this "right"; so don't trust a word that you cannot verify for yourself. Professors are notorious liars!

PROBLEM 6. REMOVAL OF CONDITIONAL MEANS. Suppose that X is an integrable random variable and $\mathcal{G}$ is a σ-field. Let $Y = E[X|G]$ and show that for $\epsilon \geq 0$ one has

$$E[(X-Y)^2 \mathbb{I}(|X-Y| \geq 2\epsilon))] \leq 4E[X^2 \mathbb{I}(|X| \geq \epsilon))].$$

Also, show that the constants 2 and 4 cannot be replaced with a smaller constants.
SUGGESTIONS: The bound is interesting even if $\mathcal{G}$ is the trivial σ-field; consider this case first. Next, you can prove the general inequality for X that just takes on two values. You can then consider X with a distribution that can be written as a finite mixture of random variables that take on just two values. Finally, show that this last result is enough to imply the inequality in general. By the way, this method of "bootstrapping" your way up from two-valued random variables is often useful, not only as a way to prove theorems but also as a way to discover theorems.

PROBLEM 7.SLIGHTLY REFINED SCHEFFÉ. Let X_n and X be non-negative. Assume that $EX_n \to EX$ and assume that for all $\epsilon > 0$ one has

$$\lim_{n\to\infty} P(X - X_n > \epsilon) = 0.$$

Show that X_n converges to X in L^1.

PROBLEM 8.CRITERION FOR CONVERGENCE IN PROBABILITY. First, review the argument from 530 that a sequence that converges in probability has a a subsequence that converges almost surely. Next, consider a sequence of random variables $\{X_n : n = 1, 2, \ldots\}$ and show that X_n converges to 0 in probability if every subsequence $n_1 < n_2 < \cdots$ has a further subsequence $m_1 < m_2 < \cdots$ such that X_{m_k} converges almost surely to 0 as k goes to infinity.

Problem 9. MC Fundamentals. Consider a Markov chain on a countable state space S. Recall the definition of ρ_{xy} and consider the relation "$\sim$" defined on pairs of states by writing $x \sim y$ if $\rho_{xy} > 0$ and $\rho_{yx} > 0$. It is useful to prove (or disprove) the following assertions for yourself:

(1) Show that "$\sim$" is indeed an equivalence relation; the first step is to look up the definition of "equivalence relation"!
(2) Show that if $x \sim y$, then x is positive recurrent if and only if y is positive recurrent.
(3) Show that if $x \sim y$, then x is null recurrent if and only if y is null recurrent.
(4) Show that if $x \sim y$, then x is has period d if and only if y is has period d.
(5) Can you think of any new "class properties", i.e. properties that hold throughout an equivalence class? You don't have to restrict your imagination to conventional properties. If you can find a weird class property, it might make a nice Toolbox essay.

Note: Look up any terms that you do not know. Make sure your proofs really are honest and logical. It is easy to lie to yourself in a problem like this.

Problem 10. Mastery of a Classic. Get yourself a copy of Erdös, Feller, Pollard (1949) from MathSciNet. Read the paper with an elementary example in mind.

(1) The authors jump into the analytical framing of the problem. Make sure you understand the probabilistic problem that is being solved.
(2) The there are two proofs in the paper. Which gives the stronger result? Is one proof particularly more general than the other?
(3) The "subsequence" proof is elementary but tricky. Read it very carefully. Be honest with yourself, and when you get to as statement that is not crystal clear, "work on it." For example, write out the "Lemma" that you think is being asserted. Now give a clear proof (or counter example!) to the Lemma that you had in mind.

Note: If you read any mathematical paper, you will find rough spots. Sorting out a rough spot is exactly the way one deepens ones skill set. Almost any paper can serve you in this way, but the paper of Erdös, Feller, Pollard (1949) is an *Über classic*; there may not be any individual paper in probability that is more famous than this — excepting some papers by Kolmogorov. This paper also motivates one to read the very beautiful paper of D.J. Newman (1975) on the Wiener theorem that is use in the first proof in Erdös, Feller, Pollard (1949).

Problem 11. Mastery of a More Modern Classic. Get your self a copy of Port (1965), "A Simple Probabilistic Proof of the Discrete Generalize Renewal Theorem". This is an instructive paper that is just some three and a half pages long. We will do the second Lemma in class, and your challenge is "just" to read the rest of the paper.

Chung once observed that it is easy to write a paper in half the usual space— just leave out every other line. It would be unfair, to say that is what happened with this paper, but it seems to me that most readers would have benefited from a somewhat longer and more patient treatment. The benefit to us, it that it offers a nice collection of exercises.

- Can you give a proof of Lemma 1? There is a nice hint in the paper as well as a reference, but you'll learn more if you make an honest attempt at the lemma before you start chasing down references.
- After you understand the set-up, the crux of the article is the arugment on pages 1296 and 1297. See if you can get to the heart of the matter by cutting out some technical details. Suppose for example that X_1 is bounded, or suppose even that X_1 has support on $\{-1, 0, 1, 2\}$. In this case the random walk is sometimes called "skip free to the left." This is often an instructive case to consider; in many ways it is the simplest relaxation from non-negative random variables to random variables that can take on at least some negative values. Such walks also show up in models; say when you bet one dollar which you can lose, but if you win your prize might be one or two dollars, etc.

Problem 12. Get yourself a copy of Aldous and Eagelson (1978) from the *Annals of Probability*. Again, this paper is brief but very instructive. It happens to be more abstract than the two papers mentioned in the preceding problems, but this abstraction is a benefit. The results are very widely applicable, yet relatively little known.

The cool thing about this paper is that it amplifies many other theorems and papers — often the authors of those theorems and papers don't know that they have some useful corollaries to their work that they have left unharvested. In particular, in many limit theorems it is semi-automatic that the "driving n" can be replace with a "random time," or that a normalization by some sequence of constants can be replaced by a sequence of random variables, or that a theorem asserting a CLT can be tweaked to give a theorem that asserts the convergence to a mixture of normals.

You may have to look up some mathematical notions to make your way through this paper, but it is worth the effort. Sorting out some of the issues that the paper suggests would make one — or more — toolbox essays.

Also, don't forget to look at some of the references. Renyi wrote brilliantly, and the ideas of stable convergence are well motivated by his original papers. You may have to make a request at the library to get the Renyi papers; they do not seem to be readily web accessible — though it is easy to find all of his papers with Erdös.

SIDEBAR: Understanding, "Rigor," and Progress.

Whenever you study a theorem you have the opportunity to go way past the basic understanding of the theorem. In the best instances, you want to "crack the bones" and "suck out the marrow." Most progress in mathematics comes out of this kind of activity — especially when combined with the creativity that it motivates.

Huge swaths of modern mathematics were driven by the desire of mathematicians to make crystal clear the ideas that were obscurely used and perhaps dimly understood by earlier mathematicians. For humdrum examples, just consider the theory of divergent series (and Euler-MacLaurin Expansions) or even the theory of analytic continuation. We see no mystery in these things now, but they were pretty spooky to people for many, many years.

The every present passage to generalization is motivated by a similar drive. After understanding a result, one inevitably asks if it is possible to extend the result; that is, one asks about a generalization. Another variation, usually more interesting to me, is to ask for a more precise result — even if the assumptions are more restrictive than those of the originals.

These are two of the basic paradigms of mathematical progress. There are several other paradigms that are worth isolation and analysis. You might want to come up with your own model of how the mysteries of the mathematical world unfold.

PROBLEM 13. GENERALIZING CONDITIONAL EXPECTATION. Let $(\Omega, \mathcal{F}, P)$ be a probability space and consider the set $\mathcal{W}$ of transformations with the following properties:

- T is a linear map defined for all integrable random variables on the space $(\Omega, \mathcal{F}, P)$.
- $T(X) \geq 0$ for all X such that $P(X \geq 0) = 1$, i.e. T is a positive transformation.
- $||T(X)||_1 \leq ||X||_1$
- $||T(X)||_\infty \leq ||X||_\infty$.

In words, T is positive and a contraction on L^1 and L^∞. For an example of such a transformation, one can take $T(X) = E[X\,|\mathcal{G}]$ where $\mathcal{G}$ is a sub-σ field of $\mathcal{F}$. Thus, the collection $\mathcal{W}$ is a super-set of the class of conditional expectations.

Let Z be an integrable random variable. Prove or disprove that the set of random variables

$$\{T(Z) : T \in \mathcal{W}\}$$

is uniformly integrable.

PROBLEM 14.NOT A SHOW STOPPER. Show that there is an increasing sequence of σ-fields $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$ and an integer valued random variable τ such that $\mathcal{F}_\tau$ is not a σ-field. Here you clearly cannot take a τ that is a stopping time. You should also recall that we have the definition

$$\mathcal{F}_\tau = \{A : A \cap \{\tau \leq n\} \in \mathcal{F}_n \quad \text{for all } n = 1, 2, \ldots\},$$

and you might do well to focus on the design of a τ such that $\mathcal{F}_\tau$ is not a λ-system.

PROBLEM 15. BOOTSTRAPPING. Suppose $\alpha > 0$ is a constant and suppose that A_n, $n = 1, 2, \ldots$ is a sequence of events such that for all $m = 1, 2, \ldots$ one has

$$\lim_{n\to\infty} P(A_n \cap A_m) = \alpha P(A_m).$$

Show that for any measurable B one has the more general relationship

$$\lim_{n\to\infty} P(A_n \cap B) = \alpha P(B).$$

Now, contemplate some interesting choices for A_n. One sequence that interests me is $A_n = \{S_n/\sqrt{n} \leq x\}$ where $S_n = X_1 + X_2 + \cdots + X_n$ and the X_i are i.i.d. with mean zero and variance 1. Now mix this up with some interesting choices for B and you get some amazing stuff.

PROBLEM 16. SPECIAL POINTS SUGGEST GENERLIZATIONS. Suppose that X is a random variable with finite mean μ and characteristic function $\phi(t)$. We know that $\mu = i\phi'(0)$ and we know that $\phi(0) = 1$. Show more generally that if t_0 is any value such that $\phi(t_0) = 1$ then we have the less well-known relation $\mu = i\phi'(t_0)$. Can you extend this observation to higher moments?

PROBLEM 17. STRATEGIC THINKING. Suppose that $F_1, F_2, \ldots, F_n$ are n continuous distribution functions with $F_i(0) = 0$ for all $1 \le i \le n$. Let X_i, $i = 1, 2, ..., n$ be independent random variables such that $F_i(x) = P(X_i \le x)$. Show that there is a stopping time τ such that

$$\frac{1}{2} E[\max_{1 \le i \le n} X_i] \le E[X_\tau].$$

To motivate yourself, think about this as a game where you are offered a sequence of "prizes" whose distributions you know but whose realizations you do not know. At time τ you say, "I'll take that prize." The result says there is a strategy that has expected pay-out that is at least half of the expected pay-out to a person who had inside information telling him the realizations of all of the prizes the day before the game.

HINT. Consider a threshold strategy where you accept X_i if X_i is the first realized prize value that is above x. You can do better than this by letting x depend on the distribution of the unseen items, but this is a start. With this (pretty stupid) rule, the probability that you accept no prize is $p = \prod F_i(x)$. Now calculate (or estimate) your expected return using this strategy in terms of x and p. Finally, pick a p and the associated x. How much better can you do with a dynamic strategy? What can you say if the X_i all have the exponential distribution?

PROBLEM 18.A MATTER OF TASTE. Prove that for integers a, b, and c one has the identity

$$\sum_i \binom{b}{i}\binom{c}{i}\binom{a+i}{b+c} = \binom{a}{b}\binom{a}{c},$$

where the sum is all over all integers i, or equivalently, over all integers for which the summands are non-zero. You can give a proof using a bijection — a truly combinatorial proof, or you can use other methods like power series or Fourier series. Why not all three?

Exercises like these are like playing the scales on a clarinet. My neighbor has been with the Philadelphia Philharmonic for 40 years and he practices scales every day.

PROBLEM 19.BORING THEN... LESS SO NOW. Suppose that X and Y are real valued random variables. Let $\mathcal{F}_0$ be the smallest σ-field such that X is $\mathcal{F}_0$ measurable. Show that Y is $\mathcal{F}_0$ measurable if and only if there is a Borel measurable function $f : \mathbb{R} \to \mathbb{R}$ such that $Y = f(X)$.

It may not even seem that this needs proof; or if it does, it might seem like the right strategy is to hurry along speaking in an authoritative voice. Nevertheless, this is a theorem with a name — it's called the Doob-Dynkin lemma.

How would you prove such a thing? Some tool is needed, such as the π-λ theorem or the monotone class lemma. To deal with this in 530 would have been boring, but at this stage of life one has a more refined sense of values. It actually is pretty interesting once you take it seriously.