Math 3070/6070 Introduction to Probability

Mon/Wed/Fri 9:00am - 9:50am

Instructor: Dr. Xiang Ji, xji4@tulane.edu

Lecture 1:Aug 18

Today

- Introduction
- Introduce yourself
- Course logistics

What is this course about?

This course will provide a calculus-based introduction to probability theory. Material covered will include fundamental axioms of probability, combinatorics, discrete and continuous random variables, multivariate distributions, expectation, and limit theorems, generally following Chapters 1-5 of the textbook. This course is a critical prerequisite for more advanced work in statistical theory and analysis.

Prerequisite

- Calculus

Why learn probability

- The subject of probability theory is the foundation upon which all of statistics is built.
- It provides you a tool to model
  - populations
  - experiments
  - almost anything else that could be considered a random phenomenon
  - example topics in Data Analysis course
- Through these models, statisticians are able to draw inferences about populations based on examination of only a part of the whole.
- A must have for any Data Scientists.

What this course WILL NOT do for you

It will not help you:

- Beat the casino at blackjack (although it may convince you that it is better not to gamble, or that a casino is a great business).
- Answer your friends' silly questions such as "What are the chances it will rain tomorrow?" (although it might make you think of ways that you might model and compute it).

Syllabus

Check course website frequently for updates and announcements.

https://tulane-math-3070.github.io/2025Fall/

HW submission

Students are required to submit hand-written homework in recitations to the TA. Homework assignments are expected every two weeks with 4-5 problems at a time.

Year 2024 comments

Your experience in this course:

- I would like to start this review off by saying that Dr. Ji is a really nice person and it is clear that he cares about his students. That being said, this course was unfortunately a huge waste of my time. Exams are structured in a way that you just need to write small enough and put all lecture notes/homework onto the allowed cheat sheet and you will get 100% without learning any material. Dr. Ji also started posting homework answers before the homework was due, eliminating any sort of motivation to go to office hours or recitation. The biggest issue I had with this class is his lecture notes. Dr. Ji simply copies the textbook, often word for word, when making his lectures. Usually, no additional explanations or proofs are given. This made learning course material extremely difficult. I feel it is highly inappropriate to attend a class in which I am simply being read a textbook by a professor – this wasted everyone's time. There were rarely example problems in lecture notes that mirrored any homework questions, making the homework unnecessarily difficult. I will end this class with a good grade having actually understood a very small amount of the content. To fix this,, the lecture notes need to be changed immediately. I know that other students have had the same complaint about this course in the past since Dr. Ji read his reviews to us at the beginning of the semester. It is unacceptable to allow this course to continue to be taught in this way.
  Response: I will let your TA post homework solutions. By the votes of other comments below and my own experience, I will respectively keep the lecture notes in the current form but encourage communications should it become a problem for anyone.

- Professor Ji is the best. He knows how to pace lectures so that everyone is on the same page. Somehow, many students don't feel comfortable speaking up and asking questions in this class. I think that's ridiculous considering the type of class environment he cultivates. The class notes are very organized and easy to follow. This class is very well designed.
- The course is difficult, but the lecture notes make studying easier.
- I think the only minor area for improvement would be that a times, the homework covered topics we had not yet gone over in class (or we went over with one day left before the homework was due) which made those questions especially difficult to complete.
- Dr. Ji is an excellent professor. Not in the research and work sense, I cannot speak to that necessarily, but in a teacher sense, he embodies what a teacher should represent.
- Great teacher I really liked the way the tests were formatted
- The lectures could be a bit dry at times. Some of the real-world applications from Dr. Ji's career made the content more engaging and I would suggest adding more in the future
- Good course, professor Ji is a good guy. I would limit the zoom option though because after the first time I didn't go back to class in person after the first time I did it. It's nice to have but ended up becoming kind of a crutch and I probably would have done better if I went in, which is my fault but I probably would have done it if I had more incentive to.
- I really loved taking the class! The professor keeps the lectures engaging, always ready to dive into questions that are being asked, and that is so key. Fostering an environment where students feel ok with asking questions is exactly what I need in a mathematics class! I would argue that grading is too easy for the class. A slightly harder grading scale (even with same type of kind, digestible tests) would be more reflective of the academic rigor of a math class at Tulane, but it is an intro class, and I am not complaining too much! :)

strongest aspects of this course

- Dr. Ji's grading system was extremely generous and we all appreciated it. This, along with recitation, was the only good aspect of this course.
- I can't stress enough how good of a professor Dr. Ji is.
- Professor Xiang Ji is very smart and helpful.
- Although this class is tough in the sense that it is really conceptual and requires you to understand a lot of derivations, I thought Dr. Ji did a very good job of setting students up to exceed through the structure of the course/grading and recognizes that the content is not easy.
- Dr. Ji, easily.

- This course is a fun learning experience, the class sessions definitely help me learn the material, and it's clear Dr. Ji wants his students to learn and tries to get them to engage.
- The cheat sheets on the exams helped me study and prepare so that I could do my best
- Professor Ji is funny asf
- 1. professor really fosters social, interractive, comfortable space. 2. you can tell prof. loves when people ask questions and is totally willing to go off-topic for a few minutes. I think that realistically, going off topic for a few minutes during class keeps me MORE engaged :) 3. Homeworks are digestible, totally not too much 4. Cheat sheets helped me study everything really well!

RateMyProfessor

- (5.0 Quality / 4.0 Difficulty) I found the course to be very challenging. Xiang is a very nice professor though.

Year 2023 comments

Your experience in this course:

- Professor Ji is awesome
- /
- There were a few times where I felt like Professor Ji's comments were a bit condescending. There were times that he would expect more from us to show proofs etc. and he would give us more time because he would assume that we were not capable enough to finish these proofs. He would then chuckle to himself, which felt condescending and really unmotivated me to go to more classes. I think there were a few instances where we also called out specific students in the class, which definitely humiliated them, but also made me uncomfortable.
- Professor Ji is a good lecturer, in my opinion. I heard many complaints that he just reads off of a pdf, but I feel his pdf is well-organized, meaning that the lecture is both easy to understand and easy to review.
- Zoom option is nice.
- Classes consist of him reading out of a textbook that is already difficult to understand just by reading. Very little is added to help you learn the material.
- It was okay at best. I did just as well on the midterm by joining the Zoom in contrast to going to class as I left the class confused 99% of the time. I taught the material to myself because Professor Ji teaches on a pdf. Managed to get a good grade because I was decent at memorizing and wrote the right things on my test cheat sheet. Ask me in a month, and I will be unable to recall anything from this class.

- Professor Ji, despite your best efforts to scare everyone away from taking this class you clearly do want us to succeeded, which I appriciate. However you could really improve the lecture time. I think that if you actually want people to come to class, you should not allow zoom. I think making attendance mandatory did not work, it just encouraged people to log into zoom. I think that if you would like people to engage more you should find a way to teach that is better than projecting a pdf. I understand that you loath writing, but reading a projected pdf is an incredibly dull way to be presented information. Even slides would be better. Writing would absolutely be the best. I think that if you want people to actually try the examples in class, you should provide clearer problems. The way you do it now is often confusing. You will show us about half of a proof or a theorem that isnt just an inequality and then say "finish this". On more than one occasion, I have found myself spending most of the two-three minutes you gave us to solve the problem just trying to figure out what you were asking. Finally, I think that you could foster a better learning enviroment. By this I mean activly encouraging questions and not laughing at your students when they ask you questions that are obvious to you. I know you are not doing this out of judgment but it does not feel great to be laughed for not understanding what you meant when you explained something the first time. Beyond this, I am a student who will ask questions no matter how to stupid I look because I think there is no other way to learn. However, I bet many other students would ask more questions if you responded in a more encouraging way.
- nice professor
- Teaching from pre-written notes rather than writing things out on the board is fine, especially since he said that was his style at the beginning of the semester. My only gripe would be that the tests were more about reciting definitions rather than applying what we learned to different problems.
- The topics in this course are pretty advanced, and I didn't understand the motivation behind a lot of it. However, the course really helped me understand where commonly used probability models come from, and that was interesting. I also really appreciated the zoom option that Dr. Ji provided for lecture, and he really seems to care about his performance as a professor. He often asked for feedback from students, and provided us the opportunity to vote on syllabus changes and class policies. He also wants his students to do well. The tests were reasonable given the course content and structure, and Dr. Ji is a lenient grader. That being said, I generally did not enjoy this course. There was so much material with extremely involved derivations and proofs that seemed unnecessarily difficult. Lectures were quite boring, and the homework assignments were really long and difficult. Sometimes it felt that Dr. Ji thought the material in this course was simple or easy, which it most definitely was not. That can be pretty discouraging for students who are feeling lost or stuck. Generally speaking though, this course was a good option to fulfill the elective requirement for my math minor and I am glad that I took it.

strongest aspects of this course

- /
- I really liked the way that the exams were organized because they were very straightforward and directly linked to what we had previously learned through the homework, classwork, etc. I also really appreciated Professor Ji's understanding to bring a one page cheatsheet, because I personally feel like memorizing a bunch of formulas is not going to help someone internalize the learning and is futile in the long-run. I think the concept of applying what you learned was a key component throughout this course, and I really liked that.
- I thought that the homework was very helpful, and I actually really liked the lecture style - I don't need the lecturer to write out each theorem or definition; it's definitely enough that Dr. Ji writes out the details of proofs and works out examples on the board.
- Exams very representative
- He went through the proofs in class. Didn't understand them but at least we were exposed to them. He also was super up-to-date with posting and updating class notes which was very convenient.
- I think the way that I truly learned the content of this class was sitting down by myself in the library with the lecture note. They did provide a comprehensive list of everything that we covered. Additionally, the exams reflected what we learned in class and while they were difficult, professor Ji was incredibly forgiving while grading. Thank you for being a kind grader.
- This is the best course I have ever taken. Textbook is awesome. Prof. Ji illustrate examples on the whiteboard very clearly. In particular, I enjoy it on zoom very much!
- He gives a clear roadmap of how to perform well in the class.
- The tests, especially the take home final, and the virtual option for lecture were the strongest aspects of this course.

RateMyProfessor

- (5.0 Quality / 3.0 Difficulty) Very nice and caring professor. Course material is not easy but he grades tests generously.
- (5.0 Quality / 3.0 Difficulty) Great professor. Very easy grader. Thinks he's really funny (he is funny, but his humor isn't for everyone). I learned a little in this class, but it was a manageable course.

Year 2022 comments

Your experience in this course:

- Exactly what I was looking for in a probability course - I got a really good grasp on the theory and this math has already come in useful in other areas and fields that I'm studying. Glad I took this class, and I appreciate the tests being more accessible and spaced out to provide less pressure - highly recommend.
- I really appreciate the lecture shift that Prof. Xiang Ji had after our midterm survey. He took suggestions seriously and dramatically improved how the content was presented to our class's needs. Having the opportunity to present and listen to classmates on related statistical topics was also fun and rewarding.
- Absolutely stupendous course. Fabulous structure, even more fabulous professor.
- Lectures were pretty disengaging. I would've rather had lecture notes written out to us rather than being read to us. Presentation extra credit opportunities were nice. Would've liked more communication and collaboration between TA and teacher.
- Once the lecture structure changed after midterms, I felt I learned much more during lectures. I think that the concepts and theories were explained clearly in class. I appreciate the generosity Professor Ji showed when grading exams, but I think that receiving more detailed feedback would have helped me learn the material better. The exams felt more like a test of our ability to make a formula sheet than a test of our understanding of course material. While this was nice in terms of my grade, I don't think this helped me with my understanding of the material. The recitations often felt disconnected from the course material because we did practice with numerical applications and calculations instead of theory. Overall I feel prepared for the second half of this course next semester.
- I felt the course provided me with very little context for why we were learning about the things we were learning. For example, now I know about a lot of different distributions and their moment generating functions, but I have no idea when I might need to use a Gamma distribution or Poisson. Additionally, I did not find the textbook to be a particularly helpful resource. There also seemed to be little communication between the professor and TA, so the activities we did in recitation weren't always relevant to what we were doing in the class.
- This course and professor were great. I got introduced to an entirely new facet of mathematics and it excited me for my major. Professor X was great and fostered a great learning environment in the classroom.
- I appreciate the professor's willingness to adjust his teaching style to include more written-out derivations. I would have preferred not to have the 10-minute presentations in class on Fridays. I felt like especially at the beginning of the semester most of the content in the presentations was more complex than what we had learned and took class time away from the material. I think it would have been helpful for the TA and Professor to communicate about the level of instruction of the course. It was hard to

understand where I stood from a knowledge perspective when the level of difficulty ranged so greatly from homework, quizzes, worksheets, and lectures. I understand that this class is a probability theory class, however, I think it would be helpful for math majors interested in applied math to have some way to learn more about the applications of probability since the Stats for scientists does not count toward the major.

- Wonderful experience! I learned a tremendous amount, and always left class wanting to know more. This course did not shy away from difficulty and I could not have appreciated it more. Far too often professors dumb down the material in order to cater to the students. Not this class, we learned intense probability theory and I could not be happier with it. I look forward to continuing my studies next semester with Statistical Inference. The rigor and complexity of the course demanded respect and, if given, the knowledge learned is powerful.
- Professor was good, but I think the subject matter was oftentimes too confusing
- The only reasons why I personally did not like the course is firstly because I do not like the subject matter and secondly, I do not like how the material was organized. I do not learn math the best when it is purely based off lecture notes. I did notice that the professor was trying to write more on the board during lectures which I did appreciate.
- I appreciate that Professor Ji mid-semester began to workout problems in class on the whiteboard as that kept me more engaged. Sometimes it was hard to follow the work on the board however as the steps didn't seem to be organized (they looked like they were written all over the place). I think it would be more easy to follow if the notes on the whiteboard were more organized linearly. Also, we only did this like once, but I think it could be a good idea to also give students problems and have them come up to the whiteboard and solve them (maybe for bonus points, doesn't have to be) as that is another way to make class more interactive. I also liked having a practice test as it always nice to get more practice.
- Class sessions were a little slow paced and repetitive for me. However, towards the end as the Professor took some feedback from students and started writing out equations on the board it was easier to follow along and I was more actively learning. I think the weekly presentations were nice but took away a chunk of class time that could've prevented us from getting behind. The weekly reviews were really helpful and I think those are great for Fridays. Our Professor, Dr. Ji, was super accommodating and very adaptive towards creating the best learning environment for students. I appreciated the mid-semester surveys and his changes based off that immensely, and his efforts to supplement our grades with bonus presentations (I just think they could be shorter or maybe uploaded on a discussion post, rather than take up 1 of every 3 classes). I liked that we had a class notes document, rather than a textbook, and this also made the class very accessible when I couldn't be there in person.
- Teaching improved significantly through the semester, he was open to feedback so that helped some. Lectures were still almost entirely him reading off of a pdf though, which did not teach me much

strongest aspects of this course

- Very good in-depth class, useful theory knowledge and strong lectures.
- The strongest aspect of this course is how it provides a very good background which is needed for future statistics courses.
- Loved the professor, absolutely no complaints. Promote Xiang Ji
- Lecture notes helpful. Loved the TA and our lab sessions.
- I really appreciated that Professor Ji asked for student feedback in the middle of the semester and adjusted the lectures based on the results. Once we started doing more derivations on the board, I was able to understand them better.
- I love you as a person, but it is seriously hard to digest anything you say in class. Reading directly from the textbook is not teaching, it is just reading. The tests do not test us on our knowledge of the material AT ALL. They simply did you write down the right thing on your cheat sheet.
- The professor and TA were very flexible when it was clear that the class didn't understand something, and were always open for feedback.
- I really enjoyed having the lecture notes typed and uploaded ahead of class. This allowed me to add to these notes and not have to get a notetaker. I also enjoyed having a notes sheet for the exams especially since there are so many formulas and distributions.
- The rigor and complexity of the course is the strongest aspect. Both Professor Zhao and Professor Ji made the material approachable and were always happy to explain and explain again the difficult concepts and proofs. I am excited to take Statistical Inference for the following semester. Additionally, the course has changed my way of thinking when assessing probabilities in all aspects of life, and there have been many instances over the semester when the knowledge imparted to me has been of service.
- The professor made himself available which I noticed and appreciated. I liked him as a person a lot and I noticed his deep knowledge on the topic.
- Switching to writing on the whiteboard vs pure lecture from typed notes. Enjoyed our bonus presentations on various math topics as it opened my eyes to how versatile statistics can be.
- I listed those above. But the professors attitude and flexibility, and use of technology to sum it up.
- professor was open to student feedback which was helpful
- I liked the presentations a lot. They were a good change of pace and way to understand how this is all applied

RateMyProfessor

- (4.0 Quality / 2.0 Difficulty) Dr Ji has a dry wit and is receptive to student feedback. He is a generous grader and offered an opportunity for generous extra credit. For the tests, he allowed a cheat sheet, and the final was take-home. I also think the tests were easy compared to how complicated they could have been. Beware the class is super theory-based similar to analysis.
- (4.0 Quality / 3.0 Difficulty) At the start of the semester I struggled with Dr. Ji's lecturing style, but after the midterm he asked for our feedback and made adjustments to his class so it was easier to follow his lessons. He didn't always explain things in great detail the first time, but if you ask questions he is always willing to clarify. Exams are also graded generously.
- (3.0 Quality / 3.0 Difficulty) Lectures are based off a pdf document which helps when you need to study, but can be terribly difficult to pay attention to in class. Tests account for about 65% of your grade but he goes pretty easy on the grading.
- (2.0 Quality / 2.0 Difficulty) Don't take this class if you're actually trying to learn the class content. I've been in here a whole semester and genuinely cannot tell you one thing I have retained. Does not communicate with the TA so recitation is not helpful either. Although, homework is graded for completion and quizzes are easy so at least its not a hard grade.
- (4.0 Quality / 5.0 Difficulty) Prof. Xi is hardcore. 3070 is definitely a theory-heavy class for people who really want to get into the underlying technical parts of probability, but if you go into it with that mindset it's really well structured and informative. Book is useful, but you have to be serious and commit time/effort into this class to do well.

Lecture 2: Aug 20

Last time

- Introduction
- Introduce yourself
- Course logistics

Today

- Continue self-introduction
- Change Midterm 1 time (conflict with Jewish High Holiday Yom Kippur)?
- Set theory (1.1)
- Axiomatic Foundations (1.2)
- Calculus of Probabilities (1.2)
- Conditional Probability (1.3)

Set Theory

One of the main objectives of a statistician is to draw conclusions about a population of objects by conducting an experiment. The first step in this endeavor is to identify the possible outcomes or, in statistical terminology, the *sample space*.

Definition The set, S, of all possible outcomes of a particular experiment is called the *sample space* for the experiment.

Example The sample space of

- tossing a coin just once, contains two outcomes, heads and tails

$$S = \{H, T\}$$

- observing reported SAT scores of randomly selected students at a certain university

$$S = \{200, 210, 220, \ldots, 780, 790, 800\}$$

- an experiment where the observation is reaction time to a certain stimulus

$$S = (0, \infty)$$

Definition An *event* is any collection of possible outcomes of an experiment, that is, any subset of S (including S itself).

Let A be an event,

- A is a subset of S,
- event A occurs if the outcome of the experiment is in the set A,
- we generally speak of the probability of an event, rather than a set.

Set operations:

- Containment:
$$A \subset B \iff x \in A \implies x \in B$$
- Equality:
$$A = B \iff A \subset B \text{ and } B \subset A$$
- Union: the union of A and B, written as $A \cup B$, is the set of elements that belong to either A or B or both
$$A \cup B = \{x : x \in A \text{ or } x \in B\}.$$
- Intersection: the intersection of A and B, written $A \cap B$, is the set of elements that belong to both A and B:
$$A \cap B = \{x : x \in A \text{ and } x \in B\}.$$
- Complementation: the complement of A, written A^c, is the set of all elements that are not in A:
$$A^c = \{x : x \notin A\}.$$

Lecture 3: Aug 22

Last time

- Set theory (1.1)

Today

- Set theory (1.1)
- Axiomatic Foundations (1.2)

Theorem For any three events, A, B, and C, defined on a sample space S,

1. Commutativity
$$A \cup B = B \cup A,$$
$$A \cap B = B \cap A;$$

2. Associativity
$$A \cup (B \cup C) = (A \cup B) \cup C,$$
$$A \cap (B \cap C) = (A \cap B) \cap C;$$

3. Distributive Laws
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C);$$

4. DeMorgan's Laws
$$(A \cup B)^c = A^c \cap B^c,$$
$$(A \cap B)^c = A^c \cup B^c;$$

We show the proof of $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ in the distributive laws. Caution: Venn diagrams are helpful in visualization, but they do not constitute a formal proof. To prove that two sets are equal, we need to show that each set contains the other.
proof:

- $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$:
Let $x \in (A \cap (B \cup C))$. By definition of intersection, $x \in (B \cup C)$ that is, either $x \in B$ or $x \in C$. Since x also must be in A, we have that either $x \in (A \cap B)$ or $x \in (A \cap C)$; therefore, $x \in ((A \cap B) \cup (A \cap C))$.
- $(A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$:
Let $x \in ((A \cap B) \cup (A \cap C))$. This implies that $x \in (A \cap B)$ or $x \in (A \cap C)$. If $x \in (A \cap B)$, then x is in both A and B. Since $x \in B$, then $x \in (B \cup C)$and thus $x \in (A \cap (B \cup C))$. It follows the same argument when $x \in (A \cap C)$, we still have $x \in (A \cap (B \cup C))$.

Definition Two events A and B are *disjoint* (or *mutually exclusive*) if $A \cap B = \varnothing$. The events $A_1, A_2, \ldots$ are *pairwise disjoint* (or *mutually exclusive*) if $A_i \cap A_j = \varnothing$ for all $i \neq j$.

Definition If $A_1, A_2, \ldots$ are pairwise disjoint and $\cup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \cdots = S$, then the collection of $A_1, A_2, \ldots$ forms a *partition* of S.

Example The sets $A_i = [i, i+1), i = 0, 1, 2, \ldots$ form a partition of $[0, \infty)$.

Basics of Probability Theory

When an experiment is performed, the realization of the experiment is an outcome in the sample space. If the experiment is performed a number of times, then

- different outcomes may occur each time
- some outcomes may repeat
- the "frequency of occurrence" of an outcome can be thought of as a probability

However, we **do not** define probabilities in terms of frequencies but instead take the mathematically simpler axiomatic approach. The axiomatic approach is not concerned with the interpretations of probabilities, but is concerned only that the probabilities are defined by a function satisfying the axioms. Interpretations of the probabilities are quite another matter:

- The "frequency of occurrence" of an event is one example of a particular interpretation of probability.
- Another possible interpretation is a subjective one, where we can think of the probability as a belief in the chance of an event occurring.

Lecture 4: Aug 25

Last time

- Set theory (1.1)

Today

- Axiomatic Foundations (1.2)

Axiomatic Foundations

For each event A in the sample space S, we want to associate with A a number between zero and one that will be called the probability of A, denoted by $\Pr(A)$. The domain of Pr is the set where the arguments of the function $\Pr(\cdot)$ are defined. It is natural to define the domain of Pr as all subsets of S, that is for each $A \subset S$, we define $\Pr(A)$ as the probability that A occurs. However, there are some technical difficulties to overcome which requires us to familiarize with the following.

Definition A collection of subsets of S is called a *sigma algebra* (or *Borel field*), denoted by $\mathcal{B}$, if it satisfies the following three properties:

1. $\varnothing \in \mathcal{B}$ (the empty set is an element of $\mathcal{B}$).
2. If $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$ ($\mathcal{B}$ is closed under complementation).
3. If $A_1, A_2, \cdots \in \mathcal{B}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$ ($\mathcal{B}$ is closed under countable unions).

From Property (1) and (2), we see that the empty set and its complement S (since $S = \varnothing^c$) are always in a sigma algebra. In fact, they construct the *trivial* algebra $\{\varnothing, S\}$ which is the smallest sigma algebra.

By DeMorgan's Law, (3) can be replaced by:

$$3'. \text{ if } A_1, A_2, \cdots \in \mathcal{B}, \text{ then } \cap_{i=1}^{\infty} A_i \in \mathcal{B}.$$

This is because:

$$(\cup_{i=1}^{\infty} A_i^c)^c = \cap_{i=1}^{\infty} A_i.$$

Example If S is finite or countable (where the elements of S can be put into one to one correspondence with a subset of the integers), then these technicalities really do not arise, for we define for a given sample space S,

$$\mathcal{B} = \{\text{all subsets of } S, \text{ including } S \text{ itself}\}.$$

If S has n elements, there are 2^n sets in $\mathcal{B}$ (why?).[hint: for each element, it is either in or out of a subset, so 2 choices].

Example Let $S = (-\infty, \infty)$, be the real line. Then $\mathcal{B}$ is chosen to contain all sets of the form

$$[a, b], (a, b], (a, b), \text{ and } [a, b)$$

for all real numbers a and b. Also, from the properties of $\mathcal{B}$, it follows that $\mathcal{B}$ contains all sets that can be formed by taking (possibly countably infinite) unions and intersections of sets of the above varieties.

We now define a probability function.

Definition Given a sample space S and an associated sigma algebra $\mathcal{B}$, a *probability function* is a function Pr with domain $\mathcal{B}$ that satisfies

1. $\Pr(A) \geqslant 0$ for all $A \in \mathcal{B}$.
2. $\Pr(S) = 1$.
3. If $A_1, A_2, \cdots \in \mathcal{B}$ are pairwise disjoint, then $\Pr(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \Pr(A_i)$.

The above three properties are usually referred to as the Axioms of Probability (or the Kolmogorov Axioms, after A. Kolmogorov, one of the fathers of probability theory). Any function that satisfies the Axioms of Probability is called a probability function.

Example Consider the simple experiment of tossing a fair coin (just once), so $S = \{H, T\}$. A reasonable probability function is the one that assigns equal probabilities to heads and tails, that is,

$$\Pr(\{H\}) = \Pr(\{T\}).$$

Since $S = \{H\} \cup \{T\}$, we have , from Axiom 2, $\Pr(\{H\} \cup \{T\}) = 1$. Also, $\{H\}$ and $\{T\}$ are disjoint, so $\Pr(\{H\} \cup \{T\}) = \Pr(\{H\}) + \Pr(\{T\})$. Collectively, we have

$$\begin{aligned} &\Pr(\{H\}) = \Pr(\{T\}) \\ &\Pr(\{H\} \cup \{T\}) = 1 \\ &\Pr(\{H\} \cup \{T\}) = \Pr(\{H\}) + \Pr(\{T\}) \end{aligned}$$

Therefore, $\Pr(\{H\}) = \Pr(\{T\}) = \frac{1}{2}$.

Example If S is finite or countable (where the elements of S can be put into $1-1$ correspondence with a subset of the integers), then these technicalities really do not arise, for we we define for a given sample space S,

$$\mathcal{B} = \{\text{all subsets of } S, \text{ including } S \text{ itself}\}.$$

If S has n elements, there are 2^n sets in $\mathcal{B}$ (why?).[hint: for each element, it is either in or out of a subset, so 2 choices].

Example Let $S = (-\infty, \infty)$, the real line. Then $\mathcal{B}$ is chosen to contain all sets of the form

$$[a, b], (a, b], (a, b), \text{ and } [a, b)$$

for all real numbers a and b. Also, from the properties of $\mathcal{B}$, it follows that $\mathcal{B}$ contains all sets that can be formed by taking (possibly countably infinite) unions and intersections of sets of the above varieties.

We now define a probability function.

Lecture 5: Aug 27

Last time

- Axiomatic Foundations (1.2)

Today

- Calculus of Probabilities (1.2)

Definition Given a sample space S and an associated sigma algebra $\mathcal{B}$, a *probability function* is a function Pr with domain $\mathcal{B}$ that satisfies

1. $\Pr(A) \geqslant 0$ for all $A \in \mathcal{B}$.
2. $\Pr(S) = 1$.
3. If $A_1, A_2, \cdots \in \mathcal{B}$ are pairwise disjoint, then $\Pr(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \Pr(A_i)$.

The above three properties are usually referred to as the Axioms of Probability (or the Kolmogorov Axioms, after A. Kolmogorov, one of the fathers of probability theory). Any function that satisfies the Axioms of Probability is called a probability function.

Example Consider the simple experiment of tossing a fair coin (just once), so $S = \{H, T\}$. A reasonable probability function is the one that assigns equal probabilities to heads and tails, that is,

$$\Pr(\{H\}) = \Pr(\{T\}).$$

Since $S = \{H\} \cup \{T\}$, we have, from Axiom 2, $\Pr(\{H\} \cup \{T\}) = 1$. Also, $\{H\}$ and $\{T\}$ are disjoint, so $\Pr(\{H\} \cup \{T\}) = \Pr(\{H\}) + \Pr(\{T\})$. Collectively, we have

$$\begin{aligned} &\Pr(\{H\}) = \Pr(\{T\}) \\ &\Pr(\{H\} \cup \{T\}) = 1 \\ &\Pr(\{H\} \cup \{T\}) = \Pr(\{H\}) + \Pr(\{T\}) \end{aligned}$$

Therefore, $\Pr(\{H\}) = \Pr(\{T\}) = \frac{1}{2}$.

Caculus of Probabilities

We start with some fairly self-evident properties of the probability function when applied to a single event.

Theorem If Pr is a probability function and A is any set in $\mathcal{B}$, then

1. $\Pr(\varnothing) = 0$, where $\varnothing$ is the empty set;
2. $\Pr(A) \leqslant 1$;
3. $\Pr(A^c) = 1 - \Pr(A)$.

proof:

- It's easy to prove (3) first. Since
  - $\Pr(A \cup A^c) = \Pr(S) = 1$,
  - A and A^c are disjoint, by axiom (3), $\Pr(A \cup A^c) = \Pr(A) + \Pr(A^c)$.

  so that $\Pr(A) + \Pr(A^c) = \Pr(S) = 1$
- with (3) proved, (1) is simple. because we know that
  - $S \cup \varnothing = S$,
  - $S \cap \varnothing = \varnothing$, they are disjoint,

  so that $\Pr(\varnothing) + \Pr(S) = \Pr(\varnothing \cup S) = \Pr(S)$.
- now for (2), $\Pr(A) = 1 - \Pr(A^c) \leqslant 1$, by axiom (1).

Lecture 6: Aug 29

Last time

- Calculus of Probabilities (1.2)

Today

- no class next Monday (Labor day)
- Redo Theorem 1 proof
- Binomial theorem
- Conditional Probability (1.3)
- Independence (1.3)

Theorem If Pr is a probability function and A and B are any sets in $\mathcal{B}$, then

1. $\Pr(B \cap A^c) = \Pr(B) - \Pr(A \cap B)$;
2. $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$;
3. If $A \subset B$, then $\Pr(A) \leqslant \Pr(B)$.

proof:

1. For (1), we have $B = \{B \cap A\} \cup \{B \cap A^c\}$ and $\{B \cap A\} \cap \{B \cap A^c\} = \varnothing$, therefore
$$\Pr(B) = \Pr(\{B \cap A\} \cup \{B \cap A^c\})$$
2. For (2), we plug in (1) first such that we only need to show $\Pr(A \cup B) = \Pr(A) + \Pr(B \cap A^c)$. Since $A \cap \{B \cap A^c\} = \varnothing$ and $A \cup B = A \cup \{B \cap A^c\}$ (use a Venn diagram, or see Exercise 1.2), we have $\Pr(A \cup B) = \Pr(A) + \Pr(B \cap A^c)$.
3. For (3), if $A \subset B$, then $A \cap B = A$. Then using (1), we have
$$0 \leqslant \Pr(B \cap A^c) = \Pr(B) - \Pr(A)$$

Formula (2) in the above theorem gives a useful inequality for the probability of an intersection (Bonferroni's Inequality):

$$\Pr(A \cap B) \geqslant \Pr(A) + \Pr(B) - 1.$$

Theorem If Pr is a probability function, then

1. $\Pr(A) = \sum_{i=1}^{\infty} \Pr(A \cap C_i)$ for any partition $C_1, C_2, \ldots$;
2. $\Pr(\cup_{i=1}^{\infty} A_i) \leqslant \sum_{i=1}^{\infty} \Pr(A_i)$ for any sets $A_1, A_2, \ldots$.

where (1) is also referred to as "Total probability" and (2) is Boole's inequality.
proof:
By definition, since $C_1, C_2, \ldots$ form a partition, we have $C_i \cap C_j = \varnothing$ for all $i \neq j$, and $S = \cup_{i=1}^{\infty} C_i$. Therefore,

$$A = A \cap S = A \cap (\cup_{i=1}^{\infty} C_i) = \cup_{i=1}^{\infty} (A \cap C_i),$$

where the last equality follows from the Distributive Law. Since $\{A \cap C_i\} \cap \{A \cap C_j\} = \varnothing$ (i.e. $A \cap C_i$ and $A \cap C_j$ are disjoint), we have

$$\Pr(A) = \Pr\left(\cup_{i=1}^{\infty} (A \cap C_i)\right) = \sum_{i=1}^{\infty} \Pr(A \cap C_i).$$

To establish Boole's Inequality, we first construct a disjoint collection $A_1^*, A_2^*, \ldots,$ with the property that $\cup_{i=1}^{\infty} A_i^* = \cup_{i=1}^{\infty} A_i$. We define A_i^* by

$$A_1^* = A_1,\ A_i^* = A_i \backslash \left(\cup_{j=1}^{i-1} A_j\right),\ i = 2, 3, \ldots,$$

where the notation $A \backslash B$ denotes the part of A that does not intersect with B. In other words, $A \backslash B = A \cap B^c$. It's easy to see that $\cup_{i=1}^{\infty} A_i^* = \cup_{i=1}^{i} nftyA_i$, and we have

$$\Pr\left(\cup_{i=1}^{\infty} A_i\right) = \Pr\left(\cup_{i=1}^{\infty} A_i^*\right) = \sum_{i=1}^{\infty} \Pr(A_i^*)$$

where the last equality holds because A_i^* are disjoint. To see this, consider any pair of $A_i^* \cap A_k^*, i > k$, then

$$\begin{aligned} A_i^* \cap A_k^* &= \left\{A_i \backslash (\cup_{j=1}^{i-1} A_j)\right\} \cap \left\{A_k \backslash (\cup_{j=1}^{k-1} A_j)\right\} \\ &= \left\{A_i \cap (\cup_{j=1}^{i-1} A_j)^c\right\} \cap \left\{A_k \cap (\cup_{j=1}^{k-1} A_j)^c\right\} \\ &= \left\{A_i \cap \left(\cap_{j=1}^{i-1} A_j^c\right)\right\} \cap \left\{A_k \cap \left(\cap_{j=1}^{k-1} A_j^c\right)\right\} \\ &= \varnothing. \end{aligned}$$

Lastly, we have $\Pr(A_i^*) \leqslant \Pr(A_i)$.

Lecture 7: Sept 3

Last time

- Conditional Probability (1.3)

Today

- Conditional Probability (1.3)
- Independence (1.3)
- Random variables

Conditional Probability

All of the probabilities that we have dealt with thus far have been unconditional probabilities. A sample space was defined and all probabilities were calculated with respect to that sample space. In many instances, however, we are in a position to update the sample space based on new information. In such cases we want to be able to update probability calculations or to calculate *conditional probabilities*.

Definition If A and B are events in S, and $\Pr(B) > 0$, then the *conditional probability* of A given B, written $\Pr(A|B)$, is

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}.$$

Note that B becomes the sample space now: $\Pr(B|B) = 1$.

Example Four cards are dealt from the top of a well-shuffled deck. What is the probability that they are the four aces? What is the probability of getting four aces at the top if knowing the first card is an ace? (there are in total 52 cards)

solution:
We define two events first. Let A be the event {4 aces on top}, and B be the event {the first card on top is an ace}. For a well-shuffled deck, all groups of 4 cards are equally likely. For the 4 aces on top, we have 4! ways of ordering (i.e., permutations of 4 distinct elements where order matters). For the rest of $52 - 4 = 48$ cards, there are 48! permutations (where again order matters). Therefore, the probability of event A is $\Pr(A) = \frac{4!48!}{52!} = \frac{4!}{52 \cdot 51 \cdot 50 \cdot 49} = \frac{1}{270{,}725}$.
Note, $\binom{n}{m}$ reads "from n choose m" (for $m \leqslant n$) and calculates by $\binom{n}{m} = \frac{n!}{(n-m)!m!}$ that gives the number of distinct combinations of choosing m elements from n total elements.
Now, let's calculate $\Pr(A|B)$. First of all, $A \subset B$, so that we have $\Pr(A \cap B) = \Pr(A)$. For $\Pr(B)$, having an ace on top instead of the other 12 kinds, $\Pr(B) = \frac{1}{13}$. Then $\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\Pr(A)}{\Pr(B)} = \frac{1}{20{,}825}$.

Lecture 8: Sept 5

Last time

- Conditional Probability (1.3)

Today

- HW2 posted
- Independence
- Random variables

Theorem (Bayes' Rule) Let $A_1, A_2, \ldots$ be a partition of the sample space, and let B be any set. Then, for each $i = 1, 2, \ldots,$

$$\Pr(A_i|B) = \frac{\Pr(B|A_i)\Pr(A_i)}{\sum_{j=1}^{\infty}\Pr(B|A_j)\Pr(A_j)}.$$

proof:
By "Total probability", we have $\Pr(B) = \sum_{j=1}^{\infty}\Pr(B \cap A_i)$ which is the denominator. Therefore, $\Pr(A_i|B) = \frac{\Pr(A_i \cap B)}{\Pr(B)} = \frac{\Pr(B|A_i)\Pr(A_i)}{\sum_{j=1}^{\infty}\Pr(B \cap A_i)}$.

Independence

Definition Two events, A and B, are *statistically independent* if

$$\Pr(A \cap B) = \Pr(A)\Pr(B)$$

Note that independence could have been defined using Bayes' rule by $\Pr(A|B) = \Pr(A)$ or $\Pr(B|A) = \Pr(B)$ as long as $\Pr(A) > 0$ or $\Pr(B) > 0$. More notation, often statisticians omit $\cap$ when writing intersection in a probability function which means $\Pr(AB) = \Pr(A \cap B)$. Sometime, statisticians use comma (,) to replace $\cap$ inside a probability function too, $\Pr(A, B) = \Pr(A \cap B)$.

Theorem If A and B are independent events, then the following pairs are also independent.

1. A and B^c,
2. A^c and B,
3. A^c and B^c.

proof:
For (1),

$$\begin{aligned}\Pr(A, B^c) &= \Pr(A) - \Pr(A, B)\\ &= \Pr(A) - \Pr(A)\Pr(B)\\ &= \Pr(A)(1 - \Pr(B))\\ &= \Pr(A)\Pr(B^c)\end{aligned}$$

For (2), we just need to switch A and B.
For (3), we have A^c and B are independent, then we can treat A^c as A' and B as B', then A' and B'^c are independent which is A^c and B^c are independent.
Alternatively, for (2),

$$\begin{aligned}\Pr(A^c, B) &= \Pr(A^c|B)\Pr(B)\\ &= [1 - \Pr(A|B)]\Pr(B)\\ &= [1 - \Pr(A)]\Pr(B)\\ &= \Pr(A^c)\Pr(B).\end{aligned}$$

And for (3),

$$\begin{aligned}\Pr(A^c, B^c) &= \Pr(A^c) - \Pr(A^c, B)\\ &= \Pr(A^c) - \Pr(A^c)\Pr(B)\\ &= \Pr(A^c)\Pr(B^c).\end{aligned}$$

Example Let the sample space S consist of the 3! permutations of the letters a, b, and c along with the three triples of each letter. Thus,

$$S = \left\{ \begin{array}{ccc} \text{aaa} & \text{bbb} & \text{ccc} \\ \text{abc} & \text{bca} & \text{cba} \\ \text{acb} & \text{bac} & \text{cab} \end{array} \right\}.$$

Furthermore, let each element of S have probability $\frac{1}{9}$. Define

$$A_i = \{i^{th} \text{ place in the triple is occupied by } a\}.$$

What are the values for $\Pr(A_i), i = 1, 2, 3$? Are they pairwise independent?
solution
It is easy to count that

$$\Pr(A_i) = \frac{1}{3}, i = 1, 2, 3,$$

and

$$\Pr(A_1, A_2) = \Pr(A_1, A_3) = \Pr(A_2, A_3) = \frac{1}{9}$$

so that A_is are pairwise independent.

Definition* A collection of events $A_1, \ldots, A_n$ are *mutually independent* if for any subcollection $A_{i_1}, \ldots, A_{i_k}$, we have

$$\Pr\left(\cap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} \Pr(A_{i_j}).$$

Random Variables

In many experiments, it is easier to deal with a summary variable than with the original probability structure.

Definition A *random variable* (r.v.) is a function from a sample space S into the real numbers.

Example In some experiments random variables are implicitly used

Examples of random variables

Experiment	Random variable
Toss two dice	X = sum of numbers
Toss a coin 25 times	X = number of heads in 25 tosses
Apply different amounts of fertilizer to corn plants	X = yield / acre

In defining a random variable, we have also defined a new sample space (the range of the random variable).

Induced probability function Suppose we have a sample space $S = \{s_1, s_2, \ldots, s_n\}$ with a probability function Pr defined on the original sample space. We define a random variable X with range $\mathcal{X} = \{x_1, \ldots, x_m\}$. We can define a probability function $\Pr_X$ on $\mathcal{X}$ in the following way. We will observe $X = x_i$ if an only if the outcome of the random experiment is an $s_j \in S$ such that $X(s_j) = x_i$. Therefore,

$$\Pr_X(X = x_i) = \Pr(\{s_j \in S : X(s_j) = x_i\}),$$

defines an *induced* probability function on $\mathcal{X}$, defined in terms of the original function Pr.

We will write $\Pr(X = x_i)$ rather than $\Pr_X(X = x_i)$ for simplicity. Note on notation: random variables will always be denoted with uppercase leeters and the realized values of the variable (or its range) will be denoted by the corresponding lowercase letters.

Example Consider the experiment of tossing a fair coin three times. Define the random variable X to be the number of heads obtained in the three tosses. A complete enumeration of the value of X for each point in the sample space is

s	HHH	HHT	HTH	THH	TTH	THT	HTT	TTT
$X(s)$	3	2	2	2	1	1	1	0

What is the range of X? What is the induced probability function $\Pr_X$?
solution:
The range for the random variable X is $\mathcal{X} = \{0, 1, 2, 3\}$. Assuming all 8 points in S has probability $\frac{1}{8}$. By simply counting, we see that the induced probability function on $\mathcal{X}$ is

x	0	1	2	3
$\Pr_X(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

So far, we have seen finite S and finite $\mathcal{X}$, and the definition of $\Pr_X$ is straightforward. If $\mathcal{X}$ is uncountable, we define the induced probability function, $\Pr_X$ for anyset $A \subset \mathcal{X}$,

$$\Pr_X(X \in A) = \Pr(\{s \in S : X(s) \in A\}).$$

This defines a legitimate probability function for which the Kolmogorov Axioms can be verified.

Distribution Functions

Distribution Functions are used to describe the behavior of a r.v.

Cumulative distribution function

Definition The *cumulative distribution function* or *cdf* of a random variable X, denoted by $F_X(x)$, is defined by

$$F_X(x) = \Pr_X(X \leqslant x), \text{ for all } x.$$

Definition The *survival function* of a random variable X, is defined by

$$S_X(x) = 1 - F_X(x) = \Pr_X(X > x).$$

Example Consider the experiment of tossing three fair coins, and let X = number of heads observed. The cdf of X is

$$F_X(x) = \begin{cases} 0 & \text{if } -\infty < x < 0 \\ \frac{1}{8} & \text{if } 0 \leqslant x < 1 \\ \frac{1}{2} & \text{if } 1 \leqslant x < 2 \\ \frac{7}{8} & \text{if } 2 \leqslant x < 3 \\ 1 & \text{if } 3 \leqslant x < \infty \end{cases}$$