IM: You know I love Chicago, but I really miss the mountains back in Colorado. Actually just nature in general. And don't tell me the lake counts. I can't hike the lake and there are always cars.

SW: Is this your version of being an annoying Texan? [laughs]

IM: Um, you know I couldn't possibly top you in that regard.

SW: No, probably not, but at least it's not my state that's relevant to today's episode.

IM: Wait, does that mean we're talking about Colorado, home of Rocky Mountain National Park, the Tattered Cover Bookstore, Casa Bonita, Garden of the Gods, and Red Rocks? The first state to legalize weed and have an openly gay male governor? THAT colorado?

SW: [laughs]

Dan Cooley: 23:30
We've seen the most extreme fires in Colorado in history have occurred in the last 20 years.

SW: This is Dan Cooley, a statistics professor at Colorado State University. And today, we'll be hearing from Dan how statisticians use extreme value theory to understand rare events. In this case, wild fires in the US and in Colorado in particular.

IM: Wait, I don't want to talk about how my home state is constantly on fire!

SW: Well that's the thing right? It's a rare event to have a major wild fire. But they're becoming less rare.

IM: And Dan is going to use extreme value theory to tell us how much more often we should expect these fires?

SW: That's the gist of it!

DC: 3:30
this is the crux of extreme value methods, and the way I like to begin to describe it is that the mantra of extreme value theory is to let the tail speak for itself.

SW: Buuuuuuut, before we get to the good stuff. Let's remind our listeners who we are. I'm Sadie Witkowski.

IM: and I'm Ian Martin

SW: And you're listening to Carry the Two, a podcast from the Institute for Mathematical and Statistical Innovation, AKA IMSI.

IM: This is the podcast where Sadie and I talk about the real world applications of mathematical and statistical research.

SW: We're the 'odd couple' of a cognitive neuroscientist and high school choir teacher, but we're still tackling these topics.

IM: And sadly, this is the last part of our collaboration with the American Geophysical Union's podcast, Third Pod from the Sun.

SW: So if you enjoy this conversation, you should check out the AGU's podcast from last friday. And don't worry, we'll make sure to link to their podcast in the show notes!

IM: Ok, now let's get to explaining some math and jargon!

SW: You're on. What shall we start with?

IM: Well I know Dan mentioned, "letting the tail speak for itself" but I'm guessing that doesn't refer to, like, an animal tail.

SW: Right. Dan was actually talking about the tail of a distribution.

DC: 2:19
so the tail of the distribution, if you picture a distribution, people probably think automatically of the normal distribution. The tail is just at the end. And so, in extreme value methods, that's what we're trying to understand and quantify, because that describes the large and often impactful observations.

IM: Ok, so then how does studying the tail, through extreme value theory, help determine the probability of extreme weather events?

DC: 3:40

perhaps the easiest way is to think of a high threshold, and you retain only the data that exceeds that threshold. And so, it seems strange at first, but we throw away 95% of the data or even a larger percentage of the data. And we only retain the data that are extreme. And then, nicely, there are probabilistic results, that it's central limit theorem like results, that say the tail of a distribution has to converge to one of these types of distributions. And the family of those is called the extreme value distributions.

SW: I'm just going to jump in and explain central limit theorem. It's a core idea in statistics that I thought a lot about when I was in grad school.

IM: Go for it.

SW: So central limit theorem basically says that in most situations with lots of samples or data points, even if you cut off the rest of the distribution and look at the tail, those values in the tail will create a distribution pattern that looks like the standard distribution.

DC: 4:40
So this gives us power to say, "Well, even though we have a very small set of data, after throwing away the bulk of the data, we still have a sensible distribution to fit." And just to speak for a moment about why we throw away all that data, if you try to fit one distribution to the entire data, well, you've got 90% of the data, a huge number of data points, that are trying to characterize the distribution. And their voice overwhelms the few data points we have out there in the tail. And we think that those few data points tell us more about the tail than something, the usual day-to-day sort of observations. And so, that's the heuristic way of explaining it, and there are mathematical justifications for thinking about it too. But we really try to let the tail speak for itself.

IM: Ok but then how do you take the data points from this chopped off tail and use them to say something about the likelihood of a future event happening?

DC: 5:38
because we use that distribution, it allows us to extrapolate further into the tail, which is challenging. If you have 50 years of data, it's challenging to say something intelligent about a 500 year event. And yet, the probabilistic results say "This is the best way we can do that." And so, we can come up with a point estimate, and we can also assess the uncertainty associated with that 500 year event. And if you're only using 50 years of data, from a single location, to estimate that, well, there's going to be a lot of uncertainty associated with that estimate.

SW: So we can use 50 years of data and try to make predictions about events that happen on the timescale of hundreds of years. But of course, there're always going to be limits to how far out we can predict with any kind of certainty.

DC: 7:20
there's always going to be a limit at which the extrapolation is practical.

IM: Of course. But let's tie this back to Colorado and wild fires.

DC 7:45
People are asking, "How does the distribution of weather, how has that changed now? How is that going to change in the future?" And that distribution of weather is really climate. And so, one of the questions that people are very interested in is, "How has the tail of that distribution changed?" And so, that brings us into extreme value methods.

SW: This ties back into some of the ideas that Richard Smith brought up when we asked him about 100-year floods and what that term actually describes.

DC 8:50
the classical language of extremes talks about a hundred year event or things like that. And that language doesn't work well for a changing climate, because the one in a hundred year event has changed in the time period. So we can talk about annual exceedance probability and the event that has a probability of being exceeded this year with one in 100. That's kind of the analog there. So I would say that extreme value theory is the best tool we have available for trying to describe the tail of the distribution.

IM: [getting impatient] Yes yes, I think I get the gist here. But let's apply it to some real weather events!

SW: [laughs] Anxious to talk about your home state huh? Alright, well let's go over an example of extreme value theory.

DC 22:31
I think we would be described as semi-arid. Fort Collins is at the intersection between the Rocky Mountains and the Great Plains. To the west of us sits high mountains. We've got mountains of greater than 13 and 14,000 feet to the west of us. Rocky Mountain National Park is essentially just west of us. The mountains are forested, pine forests. And forests burn. So this is region that has experienced fires throughout history and pre-history. The fires are a natural part of the western landscape. So one of the questions is, have the fires changed?

IM: Honestly, I don't really remember there being a lot of fires when I was growing up. Have they become more frequent since I moved or am I just more cognizant?

SW: Maybe? That really seems to be the core question we want to answer here. Have they changed, and if so, how? But this isn't a simple one-to-one question to ask.

IM: Sure, there're lots of factors that probably go into this, right?

SW: Exactly! So this is what we'd call a multivariate analysis, rather than univariate.

DC 14:41
multivariate analysis arises in a number of different settings. One is, if you have multiple variables, you can think about, with wildfires, you can think about heat and wind and dryness. And so, here, we have three variables. If they're all large at the same time, then you have extreme fire risk. And so, what we need to do is we need to describe the dependence out in the tail of this trivariate distribution. How do the extremes of dryness relate to the extremes of temperature, relate to the extremes of wind? And again, using that mantra, we're going to only use the extreme values.

SW: And even describing the extremes of wind can be complicated, like with the Santa Ana winds that create dangerous fire weather in Southern California.

DC 18:00
Essentially, there's this setup of the atmosphere, that results in, as I understand it, it essentially brings air from the desert east of Southern California into Southern California. And so, that air is warm, and it's dry. And the phenomenon that's bringing that air in, it creates high winds. And so, this can occur in the fall. And if you've got those Santa Ana conditions, then those wildfires can be very explosive.

IM: The Santa Ana winds make sense for some regions, but I didn't think that was a big factor in Colorado.

SW: You're right. For the forests of Colorado, a bigger issue might be what Dan refers to as the management signal.

DC 23:40
It's pretty clear that the management practices of the national forests have affected the nature of fire here in the West. And so, trying to piece together that story, it's a complex

story, but there are lots of factors that we believe are driving what is very clear is a change in fire behavior, recently from 50, 60 years ago.

SW: And another factor that Dan mentioned, that relates to management, is the issue of pests.

IM: Ooooo, I think I've heard of this! We're talking about the pine beetle, right? They bore into trees and end up killing whole forests if the infestation is too bad.

DC 28:00
after a beetle infestation, after you have one of these series of years where you have a large number of beetles, you end up with a lot of standing dead timber. You go to these forests, and they're these trees that are just rust colored. Because the pine needles on them have not yet fallen off, but they're dead standing timber. And you can imagine that a dead tree is going to burn differently than a live tree.

SW: So we have wind, dead trees with dry foliage, and temperature extremes that might all be playing into forest fires in Colorado.

IM: So how do researchers take all these variables and try to come up with some number to explain to the public how bad the risk is?

SW: Ah, you're talking about the fire weather index! But why don't we explain how the fire weather index is calculated and what Dan's research has to say about changing fire risks, once we're back from our break?

[music]

[ad break - why this universe]

[music ends]

IM: we're back and we're ready to talk about the fire weather index!

DC 24:37
the fire weather index is a formula, and everything that goes into that formula is a weather variable. So things like temperature and wind speed and precipitation and so forth are all input into this fire weather index. And what the formula does is it takes those weather inputs, and it breaks it into two pieces. One is the ISI, the initial spread index. The second is the BUI, the buildup index, and they're trying to get at different things.

The ISI is sort of saying, "What are the current conditions? And how ripe are they for explosive wildfires?" So current wind speed is going to be a major player in the ISI. The buildup index tries to talk about how the forest conditions are evolving over the season. And so, the amount of precipitation over the last two weeks or month very much drives the buildup index.

IM: So, ISI plus BUI equals FWI?

SW: FWI?

IM: Fire weather index. C'mon Sadie, get with the program! [laughs]

SW: I swear, I always get lost with acronyms…. Anyways! I think it's important to point out that the 'FWI' as you put it, doesn't take all possible variables into account.

DC 26:00
The fire weather index only takes in weather variables. So variables like forest condition, the fires that we saw in 2020 burned through beetle kill forests. And so, that burning through a beetle kill forest is different than burning through a relatively young forest. And so, there's no question that part of the explosiveness of the fires that we saw were due to the beetle kill. However, the fire weather index is not going to look at that, at the fact that the fire was burning through beetle kill forests. And so, this ultimately what we were interested in was, is there a climate signal? And we thought that the fire weather index, because it is a function only of weather variables, really allows us to answer the question of, how much of a climate signal are we seeing in this data?

IM: Ok so for Dan's analysis they had to cut out some variables to really hone in on their big question.

SW: Namely, is there a climate signal here and what does this say about the future for Colorado.

DC 30:17
So we did a study looking at the fire weather index, and we pulled the fire weather index from a couple of different sources. We looked at the fire weather index from re-analysis data. We looked at fire weather index from observational data. ~~W~~hat we do is we fit a statistical model to the fire weather index of seasons produced, in recent climate, basically the last 20 fire seasons, which we defined as starting in June and ending in October. So we fit a statistical model to the fire weather index, the time series of fire

weather index through the seasons. And we see a very clear signal, a strikingly clear signal here in Colorado.
31:40
When we started looking at fire weather index in Colorado, the shift in the distribution of the fire weather index is really dramatic.

IM: So basically, the answer is “yes”

SW: Yes, in that changes in climate are indeed leading to changes in the fire weather index, yeah.

DC 45:11
The way we chose to model fire was through a time series, and we thought that that was a natural approach. Because the risk of fire on June 30th very likely is dependent on the risk of fire on June 29th. We're seeing how the conditions in the forest are changing. Those conditions, again, we've worked on the fire weather index, but by keeping track of precipitation throughout the season there, the fire weather index evolves its buildup index, which talks about how the risk of fire is growing, due to a prolonged period of dryness. So there is clearly temporal dependence in fire risk, day-to-day dependence in fire risk. And so, that's why we chose a time series approach.

SW: So it’s basically a model looking at all the factors in a time-dependent fashion to try and get at whether climate changes are shifting how often rare events like wildfires happen.

IM: And this is all done at the tail ends of the distribution, *because* they’re rare events and that’s where those events are located.

SW: Exactly! And it turns out this kind of work has applications outside the typical climate modeling world.

DC 42:41
Well, extreme value methods are valuable, any time that you are trying to characterize risk associated with an unusual and large event. And so, a place where extreme value methods get used extensively is in finance and trying to assess risk associated with finance, that also carries over to insurance and other things, a reinsurance company, a company that insures the insurance companies against really extreme events. They're basically in the extreme value business.

IM: Huh, I never thought of it that way.

SW: But it makes sense right? Insurance or managing risk in your financial portfolio is all about understanding rare events.

IM: [laughs] You mean the rare event that any of us will get to retire?

SW: [sigh] In any case, I think risk and financial math is a topic that we'll be saving for a future episode of Carry the Two. Back to Dan's climate research.

DC 39:09
I think we're asking important questions, and how do we protect ourselves against really extreme events? And how does that change? How has that changed currently? How does that change in the future? I think these are important questions to be asking, and quite frankly, they're interesting statistical questions to be asking. Because there's not a one single answer. We very much have to work to improve our methods for describing the extreme phenomena arising from multiple factors. I think that's not only true in climate, that's true in finance. I think that we saw, from previous financial situations, that the models that were being used incorrectly characterized risk arising from many factors occurring at the same time.

SW: And when it comes to understanding climate change and predicting what we can expect in the future…

DC 1:02:13
I think people around the world are trying to figure out, "What is the new normal? And what do we need to do to move forward?" And these are very, very difficult questions to answer.

IM: Well good thing we have folks like Dan and all the other researchers we heard from during our collaboration with AGU to start addressing these tough questions!

[outro music starts]

SW: Don't forget to check out our show notes in the podcast description for a link to Third Pod from the Sun's story with **Dan Cooley**. We'll also link to Dan's talk on their research from the Confronting Global Climate Change program here at IMSI.

IM: And if you like the show, give us a review on apple podcast or spotify or wherever you listen. By rating and reviewing the show, you really help us spread the word about Carry the Two so that other listeners can discover us.

SW: And for more on the math research being shared at IMSI, be sure to check us out online at our homepage: IMSI dot institute. We're also on twitter at IMSI underscore institute, as well as instagram at IMSI dot institute! That's IMSI, spelled I M S I.

IM: And do you have a burning math question? Maybe you have an idea for a story on how mathematics and statistics connect with the world around us. Send us an email with your idea!

SW: You can send your feedback, ideas, and more to sadiewit AT IMSI dot institute. That's S A D I E W I T at I M S I dot institute.

IM: We'd also like to thank our audio engineer, Tyler Damme for his production on the show.

SW: And thanks to **Ty Burke**, producer with AGU's Third Pod from the Sun for their work collecting tape.

IM: And music is from Blue Dot Sessions.

SW: Lastly, Carry the Two is made possible by the Institute for Mathematical and Statistical Innovation, located on the gorgeous campus of the University of Chicago. We are supported by the National Science Foundation and the University of Chicago.