SUPER DATASCIENCE
MAKING THE COMPLEX SIMPLE

SDS PODCAST EPISODE 303: PROPER HYPOTHESIS TESTING FOR EVERY FIELD

Kirill Eremenko: This is episode number 303 with Astrophysicist and Online Data Science Instructor Sam Hinton.

Kirill Eremenko: Welcome to the SuperDataScience podcast. My name is Kirill Eremenko, Data Science Coach and Lifestyle Entrepreneur. And each week we bring you inspiring people and ideas to help you bring your successful career in data science. Thanks for being here today, and now let's make the complex simple.

Kirill Eremenko: This episode of the SuperDataScience podcast is brought to you by our very own Data Science Insider. The Data Science Insider is a weekly newsletter for data scientists, which is designed specifically to help you find out what have been the latest updates and what is the most important news in the space of data science, artificial intelligence and other technologies. It is completely free and you can sign up at superdatascience.com/dsi. And the way this works is that every week there's plenty of updates and seemingly important information coming out in the world of technology. But at the same time it is virtually impossible for a single person on a weekly basis to go through all of this and find out what is actually really relevant to a career of a data scientist and what is actually very important. And that's why our team curates the top five updates of the week, puts them into an email and sends it to you.

Kirill Eremenko: So once you sign up for the Data Science Insider, every single Friday you will receive this email in your inbox. It doesn't spam your inbox, it just arrives and has the top five updates with brief descriptions. And that's what I like the most about it, the descriptions. So you

don't actually even have to read every single article. So our team has already read these articles for you and put the summaries into the email so you can simply just read the updates in the email and be up to speed in a matter of seconds. And if you like a certain article, you can click on it and read into it further.

Kirill Eremenko: And so whether you want great ideas that can be used to boost your next project or you're just curious about the latest news in technology, the Data Science Insider is perfect for you. So once again, you can sign up at www.superdatascience.com/dsi. So make sure not to miss this opportunity and sign up for the data science insider today and that way you will join the rest of our community and start receiving the most important technology updates relevant to your career already this week.

Kirill Eremenko: Welcome back to the SuperDataScience podcast, ladies and gentlemen. Super pumped to have you on the show today, and the reason is because I just got off the phone with my friend Sam Hinton, whom I can't wait for you to meet. You know those people who are into so many different things at the same time that it just seems unbelievable that that can be possible? Well Sam is one of them, and that's why talking to him is always so fascinating.

Kirill Eremenko: So, for example, Sam, on one hand, is an astrophysicist who's doing his PhD, he's almost finished, he's got six months to go. At the same time, he was on the Survivor reality show, you know the one where you go on an island and you have to survive for months and do all these challenges. Then, again, he

was recently in Linden, which is a city in Germany, I believe, where he met 15 Nobel Prize winners, and at the same time he's into mountain biking. Then he's also launched a course on Udemy on Python for Statistical Analysis, and he knows a lot about quantum mechanics and black holes and gravitational waves and all these things.

Kirill Eremenko: So as you can imagine, our conversation today was really, really fun. So here are some of the topics that you will hear about today. We talked about meeting Nobel Prize winners, quantum mechanics, appearing on the Survivor TV show, the course that he launched in Python for Statistical Analysis. We actually went into depth on some of the topics such as hypothesis testing, we talked about academia, Python versus R, statistical significance, why p-value of 0.5 is bad, Bayesian statistics, and what is the difference between frequentist and Bayesian approaches and lots and lots more. It's a really fun podcast, I can't wait for you to check it out. Apologies right away for any background noise in my audio, I hope that doesn't affect your experience. And without further ado I bring to you Astrophysicist and Online Data Science Instructor Sam Hinton.

Kirill Eremenko: Welcome back to the SuperDataScience podcast, ladies and gentlemen. Super excited to have you on the show, because I have my good friend here with me, Sam Hinton, calling in from Brisbane. Sam, how are you going?

Sam Hinton: I'm going good, mate, how are you?

Kirill Eremenko: Amazing, amazing, thank you. And man, it's been a long time. When was the last time we caught up?

Sam Hinton: Oh, in person? Last year, the year before? You just keep traveling, it's so hard to get ahold of you.

Kirill Eremenko: You're the same, you're all over the place, it's like …

Sam Hinton: Hey, I've been back in town for a week, okay?

Kirill Eremenko: All right, we'll you're ahead of me on that one for now. Where were you? You were, what, in Canberra, in the United States, all over the world. Where was your last trip to?

Sam Hinton: Let's see, the last trip was to the Space Telescope Science Institute, which is in Baltimore in the States. I then went to the University of Philadelphia to work with a colleague there. Before that I was in Berlin, before that in Lindau, before that in Grindelwald in Switzerland, and before that I had a conference in South Africa. It's been a busy year.

Kirill Eremenko: Yeah, man. And the difference is your travel is mostly related to science, correct me if I'm wrong on that one.

Sam Hinton: Yeah, out of those trips, I did a week in Switzerland, and that was my personal travel, which was absolutely amazing. We mountain biked to the top of a few mountains, best thing ever. But everything else has been conferences, collaboration meetings, or just working with external colleagues. Lots of travel in astrophysics.

Kirill Eremenko: And what's been the most exciting one of those?

Sam Hinton: Probably Lindau, so this is the Lindau Nobel Laureates meeting. So every year there's a different set of themes and they gather as many Nobel Prize winners as they can together. So the theme for this year was cosmology, dark energy, dark matter, particle physics, and obviously that's one of my areas of research, so I got to be one of Australia's reps over there. And that was a week of banging heads with Nobel Prize winners. And it was fun. I definitely need to do better science if I ever want a Nobel Prize, those guys are absolutely insane. But definitely a highlight for me.

Kirill Eremenko: So who'd you meet from the Nobel Prize winners?

Sam Hinton: A few people. So obviously there's the Supernova team, so the 2011 Nobel Prize in physics, dark energy with Type Ia Supernovae, so Brian Schmidt and Adam Riess were there. I've met Brian before a few times, he's actually the vice-chancellor down at ANU in Australia and I've done work with him. Adam, hadn't met before. There's David Gross, so particle physics. Actually a whole bunch of people.

Sam Hinton: But the winner here, the real winner was Brian Schmidt. Because that dude got to give a public lecture in a zeppelin. Where else in your life do people say "Hey, you want to give a talk? By the way, you'll be in a zeppelin above Lindau."

Kirill Eremenko: That's crazy.

Sam Hinton: A zeppelin, like come on. I didn't get a slot in there, it was first-come first-served, and the amazing Australian internet meant by the time I logged into the system, everything was gone.

Kirill Eremenko: Oh no.

Sam Hinton: So I just stared bitterly at the zeppelin as we saw it floating around in the sky.

Kirill Eremenko: A zeppelin is like a big, one of those elongated hot air balloons with the … Well, not hot hair balloons, but the thing that flies it on, like does commercials, right?

Sam Hinton: Yeah, yeah, like a blimp.

Kirill Eremenko: Yeah, a blimp, there we go. Yeah, yeah. That's insane. Why did they do it in a zeppelin?

Sam Hinton: Because it's cool. Yeah, his lecture was probably on his area of expertise, which is astrophysics. Not much to do with a zeppelin, they actually use the zeppelin for doing measurements of the water and the coastline below, because it allows you to stay in the air for a long time actually quite cheaply, but that obviously has nothing to do with astrophysics, so it was probably just an amazing experience that any reasonable person would jump at. And maybe I'll get to do it one year, I just need to win a Nobel Prize first.

Kirill Eremenko: Well that's not far off with your rate of progress. You're doing a PhD right now, what's that in?

Sam Hinton: So the PhD, I've actually got a few topics, but I'm near the end of it so I really need to start writing a thesis properly, but instead I keep trying to get my papers out. So I'm looking at things like the large-scale structure of the universe, that is like the, I guess, imprints from the primordial universe and how they've evolved into galaxy clusters and the cosmic filament and Type Ia Supernovae, so how you can use

exploding stars as standard candles to try and characterize the expansion history of the universe, because if you can do that, you can try and determine the nature of dark energy, dark energy being essentially, and in my mind, one of the biggest unsolved questions in modern physics.

Kirill Eremenko: What is dark energy? In a nutshell, in a few sentences.

Sam Hinton: In a few sentences, okay, okay. So the universe is expanding. That expansion is accelerating, and the reason we think it is accelerating is what we call dark energy. So it's a phenomenological description here, because we don't really know what it is. The simplest explanation, this is the one put forward by Einstein before dark energy was a thing, he put forward this explanation, not for dark energy, for something else, but it actually just fits dark energy. What if spacetime itself had energy? So you take everything out of spacetime, you had a hard vacuum, but it still has energy?

Sam Hinton: And if you give it that energy, it can act essentially, it has the right density and the right pressure, because energy has both density and pressure, that's a GR thing, then you can get this sort of expansion force, this force that pushes everything away, which I guess is the wrong way of phrasing it, a force which makes space itself expand, and yeah, that is currently our best explanation of dark energy. The main issue we have is that if you try and use quantum mechanics to quantify how much energy there is in space … So you can do that in quantum mechanics, you get the wrong answer.

Sam Hinton: It's called the vacuum catastrophe. So it's not just a little bit wrong, it's really wrong. It is, in fact, the worst prediction of all time in theoretical physics. The best estimates still have you around 80 orders of magnitude wrong.

Kirill Eremenko: Wow.

Sam Hinton: Which is more wrong than estimating the size of the entire universe as less than the radius of a proton. It's incredibly wrong. And so we really need to fix that, and when I say "we," I mean people that are not me. I am not a theoretician, I am an observationalist. I will check people's models, but I really don't have the expertise to play around and actually create, delve into the nitty gritty of quantum mechanics and GR. I did quantum, I did up to like relativistic quantum mechanics in my studies, and then I was like "This stuff is nasty, let's do some more space things."

Kirill Eremenko: Yeah, I can totally relate to that. I had quantum mechanics in my fourth year, or third and fourth year at uni in my bachelor of physics. That stuff is so crazy. There's some people who understand it on an intuitive level, not me. I understood the formulas, I'm like, I can write this thing out, I can do your equations, but it's so confusing what is going on. Like, what's it called the, Heisenberg unpredictabi-

Sam Hinton: Uncertainty Principle.

Kirill Eremenko: … Uncertainty Principle?

Sam Hinton: Yeah.

Kirill Eremenko: Okay, all right. Like I can get that. But the rest of it? It's just so out of, far-fetched from this world, so different. Even probabilities are different in quantum mechanics. Crazy.

Sam Hinton: Yeah, and there's a huge ... When you get into QM, you can tell there's this huge skill gap between people that-

Kirill Eremenko: What's ... Oh, QM is quantum mechanics.

Sam Hinton: Quantum mechanics, yeah.

Kirill Eremenko: And GR is general relativity.

Sam Hinton: General relativity, yeah.

Kirill Eremenko: Okay.

Sam Hinton: But yeah, there's this huge gap between the people that know the background and know the math and can solve the equations, and those that intuitively understand them.

Kirill Eremenko: Exactly.

Sam Hinton: So I was one of the people that, I could churn through the math, that was fine. I could get the answers out at the end, but it would take me six pages of math to get there. Whilst this other person would simply look at it, draw a small Feynman diagram in like a centimeter squared, and say "The answer's probably about this."

Kirill Eremenko: Yeah.

Sam Hinton: Well, it is, but how the hell did you do that?

Kirill Eremenko: Yeah, it's crazy, man. And what I love about you is that you are so … You get into the craziest things. Like how the hell, tell me, how on earth did you get into Survivor? That was the craziest thing. When our common friend told me about this, and I watched that, there's a YouTube clip of Sam, of you-

Sam Hinton: Oh, it's so bad.

Kirill Eremenko: … It's so funny, man. I'm sorry, it's just hilarious.

Sam Hinton: Yeah.

Kirill Eremenko: How did you get, so you were-

Sam Hinton: No, it was one that got-

Kirill Eremenko: … You were on Australian Survivor, right?

Sam Hinton: Yes. Last year, so in 2018, there was the Australian Survivor Champions vs. Contenders, and somehow I was approached to be one of Australia's academic champions on the show. I have absolutely no idea how they stumbled across me. I didn't apply for the show, I didn't even know that they were casting the show, let alone, it wouldn't have occurred to me. But I was at that point in my PhD where I was starting to get a bit burnt out. I was up early in the morning, and because I work with a lot of people in the states, it meant that I would sort of finish work around 2:00 AM, because that's sort of when they're working hard in the States, because the time zones are awful.

Sam Hinton: And it was just day after day after day and I was just getting stretched real thin. And then suddenly out of the blue I get an email. I initially delete that email. It

sounds a lot like spam, right? "Hi Sam, I stumbled onto" ... Like we get a lot of these sort of emails, especially from predatory journals, so I saw their like, yeah, it's got my name right, "Do you want to apply?" And I was like, delete. And then the next week, I was like, hang on. The name was familiar, and it was the company name, Endemol Shine, and I realized now, it's because obviously I've seen them plastered, because they make The Bachelor, Survivor, My Kitchen Rules, almost every single reality TV show you've seen.

Sam Hinton: So eventually I went through and I undeleted the email, responded, had a Skype call, they flew me down to Sydney, I had a chat with the executive producer, they went to the board at Channel 10, got me all approved, and then I went to my supervisor, I was like "Hey, can I take a three month break? I really need a break." And so he's like "Where are you going?" I was like "Fiji?" "What?"

Sam Hinton: But yeah, I needed the break, and it was such a break, right? There's no technology out there, no internet, no phones. You sleep in the dirt, the show's completely real. There's no handholding behind the scenes. And yeah, it was probably one of the best experiences of my life because it was so far out of my comfort zone, but I had a blast. I think I did not too badly. I didn't win, but I wasn't first voted out, so that's something.

Kirill Eremenko: That's right, that's right man. Was it hard, doing all the social stuff? Because most of the time you're in a lab, you're working with numbers and so on, and here it is, as you say, no computers, no technology, you

have to socialize, you have to lead people, you have to be in a tribe and all these things. How different was that?

Sam Hinton: It was both easy and extremely difficult. So I tell a lot of bad jokes naturally. I don't worry or stress about too much, I just say what's on my mind and hope that it's not too stupid. And it seems like that was a good thing to do out there, because you can tell a lot of the fans of the show, they go in like "I'm going to be the biggest schemer, the most evil person of all time," and so they overdo it and you don't feel authentic. They're putting on a character, and it makes them hard to relate to and so people don't like them, they get voted out, and they're like "Where did I go wrong?" And it's like, really?

Sam Hinton: And I didn't, I just went out there to essentially get away from the PhD, relax, have some fun. And so I got on well with everyone. It was a bit hard to relate, sometimes. I was the only academic out there, I was the only unmarried person out there, I was the youngest champion. So a lot of the time the discussion revolves around kids or housing or this, and I was like "Yeah, I go to uni, cool." But that was fine.

Sam Hinton: The part I enjoyed the most was actually the challenges. I get the mental challenge a lot in my work, it's sort of why I do what I do. But the physical challenges were something that I was actually really happy with myself with. I was the only person out there to win every single one on one challenge, physical, endurance, everything. I smashed it. But I

didn't make it far enough in the show, so they didn't show most of my challenges, they edit them out.

Kirill Eremenko: Oh, interesting.

Sam Hinton: But I know I did all right, so I'm going to hold onto that.

Kirill Eremenko: Okay, all right. Wow, well what a crazy experience. And speaking of diversity of things, we talked about your PhD and travel and now Survivor, I wanted to say huge congratulations on the launch of your Python for Statistical Analysis course on Udemy and on SDS, that's really exciting, congrats man.

Sam Hinton: Yeah, thanks. That was, I think one of the biggest projects I've done so far. Went out, taking like six, seven months to develop.

Kirill Eremenko: Wow.

Sam Hinton: But I've been wanting to do it for years, because I constantly tutor at uni and encounter the same deficiencies, the same lack of practical experience. And I've pitched it to my university, "Why don't you run a course on this? Do that, make this compulsory." And the answer's always "There's no time, it's established, this is the way it's always been done." And it's like okay, well, I guess I'll make my own. And it seems to be going fairly well.

Kirill Eremenko: Incredibly well.

Sam Hinton: Obviously, yeah.

Kirill Eremenko: It's gone over, when did you launch, a few months ago, two months ago, and you've got over 1100 students,

67 ratings. You're like the highest rated course on statistics, on statistical analysis with 4.8 stars, this is incredible, man. And for your first course, especially.

Kirill Eremenko: You know what I'll tell you, first of all, design, amazing. All the pictures, images, are incredible. And then your style, you have a natural style. I watched, just when I was checking out the course I watched a couple of tutorials, just to see what's going on, and I end up, every time I end up just watching a whole section. I just can't stop, you know? It's like wow, what's he going to say next? You make jokes, you leave these cliffhangers. It's so fun to watch. I wasn't intending on watching a lot of videos at the time, I was just like, I want to check out a few. I ended up watching half an hour or 40 minutes, which is like [inaudible 00:19:13].

Sam Hinton: Stop it, you're making me blush. That was something that I know always helped me relate to students when I was giving lectures or doing tutorials was no one wants to sit and listen to just someone reciting slides, so I always keep it very conversational, very casual. If I screw up or make a mistake, I keep that in there, it's all right to make mistakes. If I have a joke, no matter how bad it is, I'll say it. I just assume that at least one person will laugh or at least breathe out just a little bit more than they normally would, and I'm happy with that.

Kirill Eremenko: Mm-hmm (affirmative), yeah, no, that's fair, fair way. Let's talk a bit about that. So I launched my course in statistics, I think, two or maybe three years ago now, or two and a half. And honestly, I haven't been looking

into it actively, so I'd love to brush up on some of this stuff. What are some of the areas you cover off in the course? What's important from the perspective of a data scientist in the space of statistics?

Sam Hinton: Right, I guess the thing that I've noticed lacking the most in the students I tutor, so again, this will have a slight astrophysics spin on it, because that's most of the people I interact with, is just being able to know what question to ask and how to answer it. A lot of the time the mathematical skills are there, people know algebra, they know basic probability theory, but being able to say "This is the question, and that implies that this is the hypothesis and this is how we'll test that hypothesis," that's where the skill gap is.

Sam Hinton: So I think the largest chapter I have is all about hypothesis-testing, and I make sure that we have a whole bunch of practical examples at the end of the course too just to try and tie everything together. So that's a big focus, and then another thing that I've noticed was really lacking is just the visual element of it. So I know when I read a textbook, if there's a page of math or a page of "This is defined as blah," and then just like eight equations after each other, it takes me so long to go through that page. I get confused so easily, I have to keep all these definitions in my mind, so I thought let's try and focus a lot on the graphical exploration of data, not just so that we understand the relationships in the data, but so that when people are making plots or figures for papers or presentations, they actually convey a lot of information very succinctly.

Sam Hinton: Because for me, if I had to choose between reading an equation and looking at a plot, I will always, always choose the plot. So there's a big focus on that as well. Obviously there's a whole bunch of refreshers on probability, Python itself, whole bunch of different chapters, extra examples, all of the good stuff to try and hammer everything home, but those are the two main things, hypothesis testing and the visual sort of workflow.

Kirill Eremenko: Okay. Interesting, so let's dive into this a bit more. So in terms of visual element, what do you use to code the visuals?

Sam Hinton: So all the plots that I use that are in the course are all done with Matplotlib, which is a Python package. We do briefly, I sort of do briefly go into how you can use other packages to get things like interactive plots or plots that you can embed in webpages if you want something a bit fancier, but Matplotlib has essentially become the standard way of plotting in the Python world, and other graphing libraries, things like Seaborn are often built on top of Matplotlib. So getting that base skill down allows you to do pretty much anything else in Python.

Kirill Eremenko: Mm-hmm (affirmative), mm-hmm (affirmative), got you. And we chatted a bit about this before the podcast, and I promised to ask this question. Python versus R. You mentioned that some of the astrophysics community actually uses R for analysis. Why'd you choose Python?

Sam Hinton: Yeah, it's simple. Because back when I was starting my research, this was a big discussion of as Python people were moving away from FORTRAN and IDL, I didn't expect anyone to have heard of IDL, it's a very proprietary astrophysics thing, but it was very expensive and awful. And so it was a big discussion, but long story short, Python won years ago. So whilst there are still people that do R and I know of some courses that are taught in R, Python is taking over. So all my students, I recommend going into that, and I'm happy to discuss why it's so much better to do Python than R if you want.

Kirill Eremenko: Yes please, tell us, it's been a big debate. Two years ago, people listening to this podcast will know, two years ago I would ask almost every single guest on the podcast, and I would call it the golden question, Python versus R, which one to pick? But lately it's become more and more transparent and obvious which way to go, and I just want to get your opinion on this.

Sam Hinton: Yeah, so the issue … A few years ago, there was debate because R had a bunch of great things like vectorized calculations that just made doing scientific analysis very natural and easy, and Python didn't quite have those. But the issue is, R still has them, but now Python has all these fantastic libraries like NumPy, SciPy, Pandas, that vectorize everything, but do so much more than you can get out of R.

Sam Hinton: So if you're, for example, wanting to go outside of academe, and this is a big thing for us, right? Only 5% of PhD students will get a full-time academic position.

There just aren't many jobs. So most of us will go into data science or software, some industry area, and if you want to, let's say, create a web app using Flask or Django, well, Python's got you covered. R? Not so much. If you want to do 100 other things, Python is popular in essentially every area of coding. R is popular in sort of data analysis. So if you want to get utility more than data analysis and scientific investigation, it's an absolute no-brainer.

Sam Hinton: And you brought this up too, if you want to do machine learning or deep learning, you want to start off with simple machine learning, well, you have Scikit-learn in Python. You want to do something a bit better but still fairly high-level, but start deep learning? Well, you load up Keras. You want to get into the nitty gritty, well you've got PyTorch and TensorFlow. It's hard for R to compete when all of these amazing packages are available for Python with minimal installation effort.

Kirill Eremenko: Yeah, no, I agree. R is conceptually different with the whole, as you said, vectorize, the way it was designed. But if I was starting out into data science now, if I was a beginner or if I was looking to even progress my skills rapidly in one of the two languages, I would no doubt pick Python. All the benefits that you mentioned, and with backing from both Google and Facebook, I think it's become an absolute no-brainer.

Sam Hinton: Yeah.

Kirill Eremenko: Yeah. All right. [crosstalk 00:26:50]

Sam Hinton: I was like, I can say more, but at that point it's just ranting.

Kirill Eremenko: Yeah. It's [inaudible 00:26:58] a bit sad. Let's hope the R guys come up with some cool stuff as well to make unique applications for R.

Sam Hinton: And then we'll watch as the Python guys just recode what they've done and it's more popular. It will happen.

Kirill Eremenko: You're giving it no chances. All right, all right. So the next thing I want to talk about is this really cool thing you mentioned, what question to ask. Indeed, that's a common issue of data scientists, that we don't ask the right questions, we dive into solving a problem, it turns out to be the wrong problem or the wrong variation of the problem or it's too long, could've been much shorter, and things like that.

Kirill Eremenko: And then you said hypothesis testing. That's like putting a kind of scientific terminology or wrapper around asking a problem. Tell us a bit about hypothesis testing. I think that mindset, that approach to, not just I'm going to ask the right question and go solve it, but I'm going to ask the right question, come up with a hypothesis, have a null hypothesis, have an alternative hypothesis, understand which one to solve. It's really powerful for data scientists because it forces you to keep statistical significance in mind. Whereas when you're just asking the question, you don't even think- unless you are used to it, you won't think about statistical significance. So tell us a bit about that. What is hypothesis testing, what's the procedure for

coming up with a hypothesis, what's a null hypothesis, what's an alternative?

Sam Hinton: Okay, let's see. This is a fairly broad topic, and obviously when you're talking about hypothesis testing, it's extremely contextual. What sort of hypothesis is entirely dependent on the question, your dataset, and what tools you have available. I'm trying to figure out where to start with this large topic. But I guess an issue people often have is coming up with a good hypothesis, so something that's quantifiable and something that has a very well-defined pathway forward.

Sam Hinton: So after, you can ask a question, then you say "Okay, now what? What do I do with this question?" And something that has always helped me is to keep in mind, what if I'm wrong? So it's very easy just to say "What if this is the case?" But if you can always say "What if this is the case, and what if it's not?" It really helps you quantify what's the difference between the two models. And obviously in this case, one of those two questions is going to be your null hypothesis, normally what if I'm wrong? What if the hypothesis isn't true and the de facto, default is at case?

Sam Hinton: And so once you try and say "What if I'm wrong?" It forces you to highlight the difference between the predictions for the null case and whatever new physics or new relationship or new idea that you're trying to see if it works with the data. So yeah, it's difficult where to go from there, because-

Kirill Eremenko: Let's go with an example. Let's, I don't know, let's take a maybe data science/astrophysics example, something off the top of your mind, something not too complex, no quantum mechanics, no GR, where you can give an example? Here's a situation, how do we ask the question, and this is the null hypothesis we would come up with and this is the alternative, this is what we want to prove, and this is in the case we're wrong. I think that would be best from here. Do you have an example?

Sam Hinton: I have a few from my work. It's simplifying them so that we can talk about it without flashing plots onscreen. How about a simple non-astrophysics example that is topical given many countries are having elections coming up is the hypothesis that an election has been rigged, or that there's election interference.

Kirill Eremenko: And you have one in the course about this, right?

Sam Hinton: Yeah, I think we have, I go over it, I think there's an extra little problem, a practical example that I run through in the section where we deal with proportion testing, which wasn't something I was going to get into in here because it involves actually putting down a bit of math which no one wants read out to them, I guarantee.

Sam Hinton: But yeah, so let's say you have potentially fishy election results. A lot of people would say, okay, "What would define fishiness? How do we know something's wrong?" And so that forces you to say "What if I'm wrong?" Or "What if there is no election interference?"

And then you have to quantify, okay, so if there is no election interference, what would we expect to see? If there is election interference, what would we now see depending on how much interference there is? Obviously any model you have will generally have a free parameter. Whether or not you actually parameterize that parameter is up to you.

Sam Hinton: I know a lot of people ... So let me quantify that a little bit. If you have election interference, you could say "What if there has been, 1 in 20 votes were changed inappropriately," a 5% sort of fraudulent fudge factor. Or you could try and say "Let's try and not fix that to 5%, and just see what does the data indicate in terms of how fraudulent do the results appear?"

Sam Hinton: And once you try and put in parameters, you can write out probabilities. Let's say, "Okay, I would expect to see this many votes out of this many people, given that I have surveyed and sampled this many people from the phone," and yeah, it's the whole process of essentially, once you have the null hypothesis and your hypothesis, you need to be able to write down the math, the probability distributions that describe that. And once you have the probability distributions that you would expect to see given your model, so that is, given no fraudulent election interference, or given X amount, how would you expect the votes to be distributed?

Sam Hinton: Once you have all of that, you then have to figure out, okay, what do I do with these probability density functions, these PDFs? And that's something we cover in the course, things like one-tailed or two-tailed tests,

how you can take your probability cutoffs and you integrate them out to actually get the chance that there was some election interference at some level of significance, because obviously this is a big thing that's often done wrong is the significance of your results, and often that gets a bit overhyped. So that was a bit of a weird explanation, forgive me.

Kirill Eremenko: No, no, all good, I was listening along. Let's see how well I understood that, let me try another one. Let's say an asteroid, speaking your language, let's say an asteroid is flying towards Earth, right? And my thing that I want to kind of prove, let's say that it's a dangerous asteroid, that it will hit Earth, that it's flying towards the planet. That would be my alternative hypothesis, that would be my H1. My null hypothesis in this case is that it's an asteroid like any other asteroid that's going to fly by, it's never going to affect Earth.

Kirill Eremenko: And so in that case I would need to describe what would I expect to see in the null hypothesis, right? So that its trajectory would not collide, would not cross Earth's path when Earth's going to in that specific place. On the other hand, in the H1, I would need to describe again what I expect to see, and I would expect to see with a high chance of certainty that its trajectory would intercept Earth's trajectory. Is that about right?

Sam Hinton: Yeah, that's sort of how you deal with it in a traditional hypothesis testing approach. The fun fact that I'll let everyone in on is that a lot of the time we don't do the traditional hypothesis testing. So especially ... Well, I

can really only speak to astrophysics, it's quite rare that you actually see someone state, "This is the null hypothesis and this is our non-null hypothesis, our test hypothesis." What we normally end up doing is actually paramaterizing both hypotheses so that they're the same thing.

Sam Hinton: So this might be, let's say we're trying to detect a signal. Well, you would have in your model parameter, a parameter that describes the signal strength. And the asteroid, when you're modeling its trajectory, you would have the parameter that might determine its initial position and velocity. Well, those are six parameters, because they're both three vectors, and then you would forward model from that point and figure out what region of parameter spaces results in disastrous impact with the planet.

Sam Hinton: And it's from those parameters that we would get our confidence intervals. So in the traditional way, you would have your hypothesis and your null hypothesis, and then you would check to see whether your hypothesis is favored using your data and to what significance it is favored. And if you, generally the way that you talk about it is that if you get more than some significance level, so the traditional one is a p-value of .05, which I recommend never to use, you would say that okay, we reject the null hypothesis.

Sam Hinton: In astrophysics, we wouldn't actually use that phrasing. We would say ... Let's stay on the asteroid example I guess. We would say that given the initial conditions and our prior uncertainty on those, the asteroid has this chance of hitting the earth. We would

just quote the number, the probability of impact, rather than actually phrasing it as "We have rejected the null hypothesis that it misses Earth."

Kirill Eremenko: Ah.

Sam Hinton: We would just say "This is the number, make of that what you will."

Kirill Eremenko: Okay. So instead of using that 95% as a threshold, you say "The chance of hitting earth is 0.1%, which is less than 5%, so we would probably reject the null hypothesis, but just for your information, it's 0.1."

Sam Hinton: Yeah, so we would use that number. We wouldn't even write about the null hypothesis. The only time that you would generally see that in astro or particle physics are on things like press releases. So if you remember a few years ago, they discovered the Higgs boson, which was a very disappointing day for all of us, because if you discover something that was predicted 40 years ago, it means that the physics is right, right? It's a validated prediction. It's always more fun when the physics is wrong, because that means there's new, undiscovered physics to go out and find.

Kirill Eremenko: Yeah.

Sam Hinton: So that's like, a lot of us would have preferred if the Higgs boson wasn't there, because then we're like "Yes, the standard model can eat the dust, we're going to find the new standard model." Turns out we can't.

Sam Hinton: Anyway, back on track. With that announcement, they saved up ... So they obviously had data for years, but they didn't publish a discovery until it hit their

required level of confidence, which for particle physics is generally five sigma. So the sort of p-value of .05 is roughly, assuming a single dimension, two sigma. Three sigma is around 99%, five sigma is you're wrong one in a million times. Not the one in 20 times that a p-value of .05 is.

Sam Hinton: And so obviously the p-value used changes with the field, but it's often just used for things like that, press releases. Because you can always publish a paper that says "Our discovery, we are 8% confident" or 3% or 1%, there's less of a focus in astrophysics on drawing that line at some given p-value and say either we have or we haven't discovered it. We prefer something with a bit more flexibility.

Kirill Eremenko: Got you, okay. From there, let's segue a little bit to statistical significance. You mentioned never to use the p-value of .05, which I'd love to hear more about. And in general, how important is statistical significance in data science? I understand it's important in academia and so on, but I'm a data scientist, I'm doing a business application, I'm doing a model, I'm providing some insights to my supervisor, and why should I care about statistical significance?

Sam Hinton: So many potential answers for that one. The very short answer is why you should care about significance is because it helps you be right. You can have a nice manager that gives you a lot of freedom, a lot of discretionary funding, and lets you just pursue whatever you want. But you have to know what it is you should be pursuing, which means you should be informing your research based upon the statistical

significance of what you've found so far. No one wants to chase down the rabbit hole for two years a finding that really wasn't that significant to begin with and turns out to be nothing but a statistical anomaly in the end.

Sam Hinton: On a different topic the way ... Well, why I would say never to trust a p-value of .05 is because it's so high. It's one in 20 times you're wrong. Now, a lot of the time if you read a paper you think, okay, well let's just say that one in 20 papers that present their results, assuming everything has an exact p-value of .05, right, is wrong. But the issue is that a paper generally doesn't present one claim. So in the process of doing your research, you often test a whole bunch of claims, or you have multiple models and you're testing different models on the data, seeing what works, what doesn't work.

Sam Hinton: Because obviously the real world is messy, and it's rare that we can actually account for every source of uncertainty, every nuisance parameter, everything we should marginalize over with perfect accuracy. And because of that we often try, okay, what if we add this parameter to account for this effect? Or what if the account for that effect has a different function or form? Whether it's just like, what if we subtract a quadratic or a cubic or a sinusoid? It's very easy to just pitch 100 different ideas and test them all very quickly. And if you restrict yourself to something like a p-value of .05, you're always going to come back with at least some model that says "Ah, okay, this one's significant, this one must be the one that works."

Sam Hinton: And so it's just so easy to mislead yourself if that's your threshold. So feel free … Well, I don't even want you to feel free to publish a p-value of .05, I think it should be much less. But just keep that in mind, and it's something that I say a whole bunch of times in the course, which is think in probabilities, not in true or false, it's significant or it's not. Because that will stop you being misled and wasting your time far more often than just "Ah yes, p-value less than .05" and it's .049 and you've tested 20 hypotheses.

Sam Hinton: XKCD has a nice infographic on this where they have a little comic where some scientists have gone out and they're testing different color jelly beans to see whether they cause cancer or some scenario like that. And it's like "Yeah, so the red one's there's no significance there. Green ones, no, yellow, no, brown, no, black, white," and they go through all of them and then suddenly it's like "Ah, purple, yes, there's a p-value of less than .05, purple jellybeans cause cancer." And obviously it's just because they've tested so many colors.

Sam Hinton: But that sort of thing happens absolutely all the time in science. And the scientists don't care too much, because we know that it happens, so when someone says the p-value of .04, we keep that in mind, we don't attribute huge significance to it, right? But once it goes out to the laypeople, once the media gets ahold of it, you get all those sensationalized rubbish titles and headlines that you probably see all the time, and that contributes a lot to I guess people's mistrust or distrust of scientific results. Because they constantly

hear "This thing here cures cancer," and it's like, well yeah, it had a 5% chance in a mouse trial. It's not something you should be getting excited about, but the news release doesn't really cover that.

Sam Hinton: So there's lots of reasons, whether it's just for your personal career to stop you wasting time, or whether it's to try and communicate accurate science, I just really detest the sort of loose science that comes from a p-value of .05. I'm not happy if one in 20 papers is wrong. That's too much. And the fact that we're testing so many different hypotheses with so many different models in a single paper means that that number is much higher. There was one study that I saw that tried to reproduce, I think several dozen psychology papers, and it could only reproduce like 40% of them.

Sam Hinton: So I remember reading about one analysis that tried to go through prior literature and just reproduce the results. And they went through dozens of claims, and they could reproduce less than half of them. So again, that's what happens when you have such a low threshold for significance, but also when you have, I guess the attitude in the publishing community, less the scientific community, that people only want to publish positive results. And this happens to me, this happens to everyone. You reproduce a result, you can't get the same thing, well, that's not really publishable by itself. You don't know whether you've made a mistake, whether they've made a mistake, whether you don't have the data, and even if you were confident in all of that, no one wants to read a study … Well, scientists don't, but the general public doesn't even

more, want to read a study that just says "We couldn't reproduce this." It's not new, it's not interesting or novel, so it's a lot harder to get published and it's even harder to get funding for.

Sam Hinton: I have never heard of anyone getting funding to just go back and try and double-check a whole bunch of results. Only if you're improving in some way or doing something different does anyone decide to give you money. And I can half see why, right? We prefer new and interesting novel discoveries. But sometimes you just have to suck it up, do the nitty gritty, and make sure that what's been put out there already is actually accurate.

Kirill Eremenko: It almost feels like there should be a system where whenever a paper gets published there's equivalent funding set out for somebody else independently to reproduce that same result, just to make sure that what is being published is not erroneous.

Sam Hinton: Yeah, that would be absolutely fantastic, and if you can get that set up, that would be great.

Kirill Eremenko: Yeah, let me-

Sam Hinton: I don't have the political clout to do so.

Kirill Eremenko: … Let me get that pot of gold that's under my bed.

Sam Hinton: Yeah, that's right. It's something that we're actually trying to, we're very good with this, actually, in astrophysics, and it's one of the areas that I like most about this community. So when we did an analysis of gravitational weak lensing in the dark energy survey,

we did multiple models, and they were blinded so that we didn't know the end result.

Sam Hinton: So we developed two independent models, run them both over the data set in exactly the same way, get the blind answers out at the end, so we've added fudge factors and scaled our results, we don't know what the true value is, and then at a live teleconference in front of everyone, you unblind both of the methods. And ideally, they should agree within their uncertainty to some value that is physically reasonable. And that's happened so far for us, which is good. But yeah, that sort of having multiple methods that are consistent with each other is something that is a lot easier for us to do in astrophysics, because we have data products that exist, and is harder for people to do, let's say in psychology, where you don't particularly have, the issue's not with the model, the issue's with your confounding factors and your survey sizes, so it's not like you can just have slightly different physics that go in. You don't deal with physics, you just deal with human randomness which is unfortunately impossible to model.

Kirill Eremenko: Yep, yeah. No, I'm totally with you. I was actually reading, I read an article maybe a year or actually two years ago, I think it was on Science or Nature magazine online, which was talking about exactly that, that p-values are causing a lot of problematic research to be published or research that's misleading to be published, and they also tried this reproducibility of results and [inaudible 00:48:56] conclusions, and yeah, one example, they're saying that you can

actually use p-values, and if you collect stamp data you can prove that frogs can predict earthquakes, random things like that.

Kirill Eremenko: And this leads me to another topic I want to talk to you about, which I think you're quite fond of, and that is Bayesian statistics. What is the difference between frequentist statistics, that's where we use the p-values and that was developed by Fisher in the early 20th century, and which is taught in schools, which is the norm in the scientific community. What is the difference between that type of statistics and Bayesian statistics?

Sam Hinton: Okay. There are a few, and often the differences that you care about are contextual. So if you're implementing a model, if you have some model that you're coding up, the difference between, let's say a frequentist and a Bayesian approach are simply the priors. So in a Bayesian framework, you take your prior information that you have on, let's say the physical distribution of your model parameters into account.

Sam Hinton: Another way of thinking about it on a more conceptual level is that I guess Bayesian statistics, your model parameter, so let's say you're fitting a line, your model parameter might be your gradient and your y-intercept, right? Your model parameters are unknown, and your data is what's fixed. And that's how you sort of conduct your analysis. In the frequentist way, the data is what's unknown, but the model parameters are fixed. And so you're sort of asking the reverse question, what's the probability of getting your data

given your model parameters, or what's the probability of your model parameters given your data? And obviously the difference between those is how you-

Kirill Eremenko: Sorry, could you repeat that? I think that flew over my head, could you repeat that again please?

Sam Hinton: Right, okay, so one of the key differences is just the order in which you look at things.

Kirill Eremenko: Okay.

Sam Hinton: In that, for, so you might have the probability of your model parameters given your data, and that's sort of a Bayesian approach, or the probability of your data given your model parameters.

Kirill Eremenko: Ah, okay.

Sam Hinton: So it's like which one is fixed and which one is the random variable? So in Bayesian statistics our data's fixed and our model is the random variable, versus I guess the opposite. But in terms of implementation details, both of those methods when you actually implement a model need a likelihood, where a likelihood is what's the probability of your data given your model parameters.

Sam Hinton: The difference is that when you add in your prior, so what's the probability of our model parameters not caring about the data, what's our prior information or our prior knowledge on those parameters, well that's how you can unite the two through something called Bayes Theorem.

Sam Hinton: Now, the primary benefit, I would say, of Bayesian statistics, is that if you actually write out Bayes Theorem in full, so it's the likelihood times the prior, we just talked about, but then it's divided by an integral over the data, and that's called the evidence. Now unlike frequentist statistics, Bayesian statistics has a very nice way of comparing different models to each other. You compute the evidence for each model and you can pair those, you have an evidence ratio, and you can use that ratio to say "Hey, given these two models, this one is preferred by this much."

Sam Hinton: Now there are analogous ways of doing this in frequentist statistics. You can, let's say you have a traditional chi-squared approach, a chi-squared approach being where you simply have my data is this vector, my predictions are this other vector, I take the difference between them, so how close am I, and then you divide that by your error. That's a chi-squared approach.

Sam Hinton: And there are ways of approximating model selection, AIC, BIC, DIC, a whole bunch of ... the IC being information criterion, different ways of comparing models. But they're not as good. They're simply not as statistically robust as something like the Bayesian evidence. The reason, of course, that Bayesian statistics is now getting big whilst frequentist statistics is sort of seen as the older traditional way of doing things is simply computational power. Computing the evidence is an absolute nightmare, and only with modern computers do we have any chance of doing that.

Kirill Eremenko: So before we weren't able to compute the evidence, now we are.

Sam Hinton: Yeah, so before you sort of restricted to let's say a one-dimensional or a two-dimensional problem. It's essentially an integral over every parameter. So if you had something like, sort of my last supernova cosmology model had 1000 parameters. It's very numerically difficult to compute the integral in a 1000-dimensional space. It requires sophisticated numerical techniques, essentially something called nested sampling, which we probably don't have time to get into, but we can do that these days, we couldn't in the past.

Kirill Eremenko: Okay, got you. But there has to be a view in the scientific community or data science community that Bayesian statistics is more correct or is more powerful than frequentist statistics. The computational power, yeah, great, it's there now, so we can do Bayesian statistics, but people wouldn't be running or moving towards it if they didn't see that it's more valuable than frequentist, so ...

Sam Hinton: Oh, for sure. It's definitely ... Yeah, so it's more valuable because it is both more correct and more robust. So as we were talking about before with, for example, the evidence ratio, so your Bayesian evidence, that's a much better way to try and discriminate between multiple models than the approximations that we would use with frequentist statistics.

Sam Hinton: The other thing is that Bayesian statistics allows us to better make use of our actual knowledge, so when you include things like the priors, you inform your final answer based not just upon your data but what you already knew before. So conceptually, imagine if you had two studies. You did the first study, you go into that study knowing nothing. You finish the study, and you have some … It wasn't that good a study, a first pass, very rough, very interim, but it gives you some information on your model parameters.

Sam Hinton: And you can use that information as the prior in the second study, and that way you can more rigorously combine your knowledge of your model parameters going forward. The way that you would try and do that under frequentist statistics is you would have two separate studies, they would give you model constraints, so X is this number plus or minus this number. You would then say, okay, we're going to assume that these areas are both Gaussian, and we're going to combine them together, and so Gaussian error propagation.

Sam Hinton: Which is fine, if things are Gaussian. But in the real world, lots of things are not Gaussian, lots of things are not even close. So we really need to try, especially once you get into precise statistics, so when you're constraining model parameters down to sub-percent levels, you need to make sure that you've done it correctly, otherwise you end up having systematic biases in your model constraints.

Kirill Eremenko: Amazing, amazing. I think that's more than enough to process right now, thank you for the overview. So in

terms of business, what would you recommend for people? Is it worth looking into Bayesian statistics now, or start with frequentist statistics?

Sam Hinton: Oh God no. There's no point starting with frequentist now, because it's sort of like, do you want to learn to paint, should you first learn first with black and white or with color?

Kirill Eremenko: Color.

Sam Hinton: And it's like yeah, you just start with color. You would learn composition and tonality and shading all in one go instead of trying to piece it together, and you get a much more cohesive view if you do just start with everything, because you go into it with the right concepts. So if you do [crosstalk 00:57:45]-

Kirill Eremenko: But Sam, it sounds so hard. Bayesian statistics to frequentist statistics, you just go and you're like, p-values, okay, I can handle that. But Bayesian, there's the priors, there's the evidence, then there's the reverse relationship between the probability of parameters and the data, things like that. It just sounds really complex. Is there an easy way to learn Bayesian statistics?

Sam Hinton: Yeah, just do it. So obviously there's a terminology overhead, but the concepts are all quite simple, and the probability that goes in is actually not that complicated. It's something that you can do just with the very basic probability identities. A lot of the time the application of Bayesian statistics can get complicated, but that's not because anything Bayesian is complicated, it's because your likelihood [inaudible

00:58:40]. And that's something that is both in frequentist and Bayesian statistics.

Sam Hinton: So once you start getting into the nitty gritty, they're just as complicated as each other, it's just have you spent the extra few minutes at the start trying to learn what these terms mean? And I know that we have, so for new students in cosmology, one of my colleagues I know has helped write a book that's called Bayesian Methods in Cosmology, and it starts with a whole bunch of astrophysical examples to show you exactly how you should formulate things, and once people do one or two of those, it just becomes natural, because Bayesian thinking is how we naturally think.

Sam Hinton: There's another XKCD comic that says "Here's a neutrino detector. It will roll two dice, and if it gets two sixes, it lies to us." Right? And then they keep this detector on, and at some point the detector says "We have detected a neutrino burst, the sun must have gone supernova." And the frequentist would say, okay, well, let's see, one in 36, that's less than .05, so we have detected the sun going supernova, which doesn't make sense to us conceptually, right? But a Bayesian would say, well, I have a pretty good prior, so my prior knowledge that the sun hasn't gon supernova in billions of years, so why would it go supernova now? And so the Bayesian statistician would say "Actually, really don't think it has."

Kirill Eremenko: Wow, that's a fantastic example of priors. Indeed, when you were talking about priors I was thinking what example could we give? That's beautiful. Frequentists would look at that, "Okay, less than .05,

that's it, must be statistically significant, the sun has exploded." On the other hand, Bayesian is like not only do they look at the evidence now, they look at the prior knowledge. That's so cool.

Sam Hinton: Yeah.

Kirill Eremenko: Fantastic, okay. Very cool. And I think that puts it really into perspective. Okay. Well, Sam, been really nice chatting to you. We've almost depleted our time for this podcast, even though there's so many more things I would love to ask you. So just quickly, what's your research coming up now, like how much more are you doing your PhD for?

Sam Hinton: Well hopefully I'll have wrapped up the PhD in the next six months, and then I'm looking to try and start working with DESI, the dark-energy spectroscopic instrument, so they're going to survey millions upon millions of galaxies in the night sky to try and map out the imprints from the early universe in the galaxy distributions. They'll do some absolutely amazing science, so I'm waiting on all the government funding bodies to get back with either "You've been accepted; rejected", or "This is how many people you can hire," and I'm going to try and jump into whatever good position I can so that I can set myself up for the next few years doing some good physics.

Kirill Eremenko: Okay, what's your dream position that you would love to get?

Sam Hinton: Oh, well that would probably be $1 million a year, stay at home, and do whatever I want. Failing that, I would really love just a stable, so the big issue in academia is

that often you have very short contracts, one, two, or three years and then you move country every couple of years. Which is fine until, let's say, you want to get a pet, or you have a partner or you want to buy a house, right? Then it gets in the way. So a stable job somewhere, probably in Australia, because I happen to have friends and family, that's split between doing good research and good scientific outreach, because I do do a lot of public talking, I try and get out there, spread the good word of science to everyone. And if I can get paid to do that instead of just volunteering, that would be amazing as well.

Kirill Eremenko: Fantastic, fantastic. Well, if anybody listening to this podcast is in the scientific community, especially in Australia, and you happen to be working in astrophysics, get in touch with Sam.

Sam Hinton: There'll be hundreds of them, I'm sure.

Kirill Eremenko: Sam, speaking of that, what's the best way to get in touch with you?

Sam Hinton: I would say flick me an email. That's obviously probably the best way. I do get a lot of emails, though, so if you want to get something less formal, I have public profiles on the usual, Instagram, Twitter, or LinkedIn that you can contact me there. I don't know when I would get back to those, hopefully within a couple days, and obviously I have my website up, cosmiccoding.com.au, that details some of the projects I've been in, some of the stuff I've been doing, and other various ways of getting in touch with me, so that's probably a good resource if you have any

questions space-related and you want them answered, happy to help out.

Kirill Eremenko: Fantastic, and we'll include all those links in the show notes if somebody wants to get in touch, and I'd encourage everybody to get in touch. And before I let you go, one final question, what's a book you can recommend to our listeners to help them on their journey into statistics?

Sam Hinton: Oh, okay. Well, the book that I would have to recommend, I talked about it a little bit before, Bayesian Methods in Cosmology. It's partially, written by a friend of mine, a very fantastic overview of Bayesian statistics and fun applications in how they can be used in space science. Obviously I'm a bit biased, I like space, I like space science. Even if you don't, it is a wonderful introduction to Bayesian methods.

Kirill Eremenko: Got you. There we go, Bayesian Methods in Cosmology. Who's the author?

Sam Hinton: Oh, there's a whole bunch of them. We should probably put a link down there, because this is common in academic works, every single chapter has a different author.

Kirill Eremenko: Oh wow. I like books like that, it gives you a lot of different perspectives on things.

Sam Hinton: Yeah.

Kirill Eremenko: That's fun. Okay, so once again, thanks so much my friend, I hope we catch up sometime soon, and once again, congrats on your epic course launch. We'll

include the links to that in the show notes if anybody's interested, and yeah, best of luck with your PhD. Sounds like exciting times, you're finishing it up.

Sam Hinton: I hope so. Well thanks for having me, mate.

Kirill Eremenko: Thank you ladies and gentlemen for being on the SuperDataScience podcast today. I'm super excited that you got to meet my friend Sam, and I really hope you enjoyed all the insights that he shared with you today. Probably the biggest takeaway for me from this episode, everything was really fun, really exciting. But the thing that really stuck in my mind is, that great example about Bayesian inference, about the sun and how the sun hasn't exploded yet, so we're using that prior knowledge in our statistics calculation, probability calculations, and that informs better our assessment of the current situation and our predictions for the future.

Kirill Eremenko: And look out for opportunities like that in business situations, and maybe we'll do a podcast later on, we'll definitely do a podcast later on on how to apply that better in business cases, and that just shows the power of Bayesian inference and how it's different to frequentist statistics.

Kirill Eremenko: On that note, as always, you can get the show notes for this episode at superdatascience.com/303, that's superdatascience.com/303. There you can find the transcript for this episode, a URL for Sam's LinkedIn, and all the materials we mentioned on this episode, including a special coupon link to Sam's course on Udemy. So if you're part of SuperDataScience, the

membership, then you already have this course included in your membership, and you can access it on the SuperDataScience website. But if you want to get access to this individual course by itself and you're not part of the SuperDataScience membership, you can find a special coupon in the show notes at superdatascience.com/303.

Kirill Eremenko: And yeah, if you enjoy this episode, then share it with your friends, don't just keep it to yourself. Maybe you have an astrophysicist who you know or somebody interested in astrophysics, or somebody interested in statistics and Python and learning all these amazing things, or somebody who you think might resonate with Sam's personality. They're easy to share, send them a link to the episode, superdatascience.com/303 and they can get onboard with all these great insights.

Kirill Eremenko: And on that note, thank you so much for being here today. I look forward to seeing you back here next time, and until then, happy analyzing.