Episode Book

EPISODE #6459

Why the Gaussian Distribution Fails in Plain Sight

PUBLISHED	5 November 2026 at 02:31 UTC
DURATION	30:32
LISTEN	myweirdprompts.com/episode/gaussian-distribution-thin-tails-assumptions-82965/
HOSTS	Corn, Herman, Hilbert
SCRIPT MODEL	z-ai/glm-5.3
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.26
SOURCES	23 cited by the research (listed at the end)
BOOK GENERATED	5 November 2026 at 07:03 UTC

SYNOPSIS

The Gaussian distribution underpins everything from least-squares regression to modern AI diffusion models, controlled entirely by just two parameters: mean and variance. But its aggressive, thin-tailed exponential decay makes extreme real-world events vanish mathematically. Exploring the mathematical power of the central limit theorem alongside the subtle dangers of confusing bell curves with heavy-tailed distributions reveals why treating multiplicative processes as additive leads to impossible conclusions like negative hospital stays.

TRANSCRIPT



Corn

Seventy-eight percent. That's how much of the kurtosis in ten thousand US stock market observations came from a single trading day. One day. Out of ten thousand.



Herman

One observation carrying most of the shape of the entire dataset, and if you'd modeled those returns with a Gaussian, that day was practically impossible. Which is a strange way to open an episode about a curve, but it's exactly where Daniel wants us to go.



Corn

Daniel's question is the opposite of exotic. He wants to know what a Gaussian distribution actually is, precisely, because the term keeps turning up, including in that episode we did on synthetic data pitfalls, and he suspects he's been nodding along. His words, roughly, are "for the unenlightened masses, myself included." He wants the real statistical definition and then what it does in the world, in AI engineering and everywhere else.



Herman

And it's a good time for it, because the Gaussian is quietly one of the load-bearing assumptions of the entire AI stack. Diffusion models, the things generating half the images on the internet right now, are literally built out of Gaussian noise.



Corn

So let's start with the definition, because the definition is where the trouble starts.



Herman

A Gaussian distribution, also called the normal distribution, is a continuous probability distribution over the real line, and here's the remarkable thing about it. It has exactly two knobs. There's mu, the mean, which slides the whole thing left or right, and sigma squared, the variance, which stretches or squeezes it. That's the entire family. If you tell me the mean and the variance of a Gaussian, I know everything there is to know about it. There's no third parameter, no skew knob, no tail knob, nothing.



Corn

Two knobs, forever.



Herman

The standard normal is the one with mean zero and variance one, and its density at the peak, right at zero, is one over the square root of two pi, which is about zero point four. The full density function is one over the square root of two pi sigma squared, times e to the power of minus x minus mu, squared, over two sigma squared.



Corn

I want to flag something for the unenlightened masses, as Daniel calls them. That formula is not just decoration. The e to the minus something squared part is where all the character lives. The density falls off as the square of the distance from the mean, inside an exponential. Squared distance, then exponential decay. That's a specific and, as it turns out, aggressive way to die off.



Herman

Aggressive is the right word. The Gaussian has what people call thin tails or light tails. Once you get a few standard deviations out, the probability doesn't just get small, it gets small at a rate that makes extreme events almost vanish. And there's a detail I love, which is that the cumulative distribution function, the thing that tells you the probability of landing below some value, has no closed form. You can write the density on a napkin, but the area under it can't be expressed in elementary functions. It's computed numerically, through something called the error function. Every statistics textbook in the world is carrying around a table because of that.



Corn

So the curve is elementary and the area under it isn't.



Herman

Yes, and that's a good reminder that "convenient" in this context means convenient in a very particular nineteenth-century way.



Corn

Which brings me to the bell curve thing, because this is where I think casual conversation goes wrong first. People say bell curve as if it's a synonym for Gaussian. But you told me once, and I checked and it's true, that plenty of distributions are bell-shaped. The Cauchy is bell-shaped. Student's t is bell-shaped.



Herman

The logistic distribution is bell-shaped. And here's the kicker: the Cauchy looks so much like a Gaussian at a glance that if you histogram data from one, most people would happily call it normal. But the Cauchy doesn't even have a well-defined mean. Its tails are so heavy that averaging more samples doesn't converge to anything. You can't tell them apart in the middle. The entire difference between a Gaussian and a Cauchy lives in the tails, which is exactly the part of the histogram you can't see, because tail events are by definition rare. You'd need enormous amounts of data before the corners of your histogram start telling the truth.



Corn

And the part you can't see is the part that decides whether your model survives contact with reality.



Herman

That's the tension I'd want every listener to hold onto for the next twenty minutes. The Gaussian is simultaneously the best-justified distribution in statistics and, in a lot of the situations where it's actually used, quietly wrong.



Corn

So justify it. Why does this thing keep showing up? Because it does. Height, measurement error, exam scores. If you've seen one histogram in your life, you've seen something bell-ish.



Herman

The main engine is the central limit theorem, and it's one of the astonishing results in mathematics. Take any random variable with a finite mean and a finite variance. It can be ugly, it can be lopsided, it can be discrete, doesn't matter. Now take the average of many independent draws from it. As the number of draws grows, the distribution of that average converges to a Gaussian. Not approximately, not usually. It converges, full stop.



Corn

The lumpy thing averages itself into a bell.



Herman

That's why measurement error is the canonical example. When you weigh something on a scale, your error is the sum of a hundred tiny independent perturbations. Air currents, vibration, the slight unevenness of the sensor, your hand not being perfectly still. Each of those is some arbitrary random variable, and their sum is dragged toward the Gaussian by the theorem. Any physical quantity that is fundamentally a sum of many small independent effects tends to look nearly normal. That's not mysticism, that's the theorem doing its work.



Corn

And it's a real justification. It's not just that statisticians like the shape.



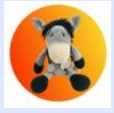
Herman

It's a theorem! It's one of the few places where the assumption is actually earned. If your quantity is a sum of many independent contributions with finite variance, the Gaussian is the right answer, provably. But notice the two conditions hiding in that sentence, because both of them fail in the wild all the time. Independent. Finite variance. Financial returns are famously neither. They cluster and they have tails that make variance estimates unstable.



Corn

You said the central limit theorem is the main engine. Which means there's a second engine.



Herman

Convenience, and I don't mean that dismissively. The Gaussian is closed under linear combination. If you take independent normal variables and add them, multiply them by constants, take linear combinations, the result is normal again. Every time. That's why propagation of uncertainty works with tidy formulas, why least-squares fitting, which is the workhorse of all of statistics and a large fraction of machine learning, has closed-form solutions. The entire machinery falls out of that one property.



Corn

Mean squared error is a Gaussian wearing a trench coat.



Herman

In a meaningful sense, yes. When you train a model by minimizing squared error, you're implicitly assuming the noise around your prediction is Gaussian, whether you said so or not. Nobody in a machine learning pipeline announces "I am assuming normality." They just reach for the loss function that has a nice gradient, and the assumption arrives in the same box.



Corn

I find that slightly sinister. The assumption doesn't get made. It gets inherited.



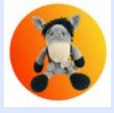
Herman

And there's a cousin that I have to mention, because confusing it with the Gaussian is one of the most common modeling errors in practice. The log-normal. If X is log-normal, then the natural log of X is normal. And the rule of thumb is beautiful: sums go Gaussian, products go log-normal. Anything that's the sum of many independent effects goes normal by the central limit theorem. Anything that's the product of many independent positive factors, energies, concentrations, particle sizes, prices over time, goes log-normal, because the theorem applies in the log domain. You take the log, get a sum, and the theorem kicks in.



Corn

So a stock price over time is multiplicative. Each day's return multiplies the previous price.



Herman

Right, and that's the classical argument for why prices should be log-normal, which is the assumption sitting under Black-Scholes and a huge amount of twentieth-century finance. Notice a log-normal variable can never be negative, which fits prices, and it's skewed, right-tailed, which also fits. It's a better model than a plain Gaussian for prices. Whether it's good enough is the question that, as we'll get to later, Nassim Taleb has spent a career answering with the word no.



Corn

Before we get to the fighting, I want the misuse warning stated plainly, because the reference literature states it plainly and it deserves saying out loud. Normal distributions are frequently used where the data are not normal, and in those contexts the normal distribution is a poor model. That's not a fringe complaint. That's the baseline entry on the subject.



Herman

It's almost comical how blunt it is. The standard reference on the normal distribution contains, in effect, a warning label against its own most common use. And the misuse has a shape to it. People look at the middle of their histogram, it looks like a bell, they declare normality, and everything downstream, their confidence intervals, their risk figures, their five-sigma thresholds, inherits an assumption about the tails that was never checked. Because, again, the tails are the part you can't see.



Corn

Herman, let me ask you the practical version of this. You spent decades in medicine. You actually ran the trials. Where did you see the assumption hold and where did you see it rot?



Herman

Oh, all over, both directions. Lab measurement error is the honest case. Run an assay on the same sample a hundred times and you get a beautiful bell, because the errors really are a sum of many small independent perturbations, and the central limit theorem is doing exactly what it says. But the moment you're dealing with a biological quantity that's the product of processes rather than the sum, enzyme concentrations, hospital length of stay, people's incomes if you're doing health economics, you get skew and heavy right tails, and treating those as normal produces nonsense confidence intervals on the low end. A normal model of hospital stay will cheerfully tell you some patients stay minus two days.



Corn

Negative days in the ward.



Herman

Physically impossible, statistically permitted. The distribution didn't care. That's my rule of thumb for listeners: if the quantity can't go negative and it's built multiplicatively, the Gaussian is the wrong family before you've collected a single data point.



Hilbert

Hilbert's Anteater Trivia. Giant anteaters range from Central America down to northern Argentina. I have also ranged widely, and nobody has yet produced a map.



Corn

Nobody's produced a map because nobody asked.



Herman

Moving on before Corn develops a theory about that. Where was I. Products versus sums. So that's the statistics, Corn, that's the foundation. Two knobs, squared-distance exponential decay, thin tails, the central limit theorem as justification, convenience as the second engine. Now let's talk about where this actually bites, and I'd start inside AI, because the Gaussian isn't just assumed there. It's installed.



Corn

Explain that, because I think it would surprise people. When we did the episode on how transformers came to dominate everything, nobody mentioned that the shape of a distribution was a design decision.



Herman

Because it usually isn't presented as one! But diffusion models, the generative architecture behind most modern image generation, are built on the Gaussian in the most literal way. The forward process takes your image and progressively destroys it by adding Gaussian noise, sample by sample, until the image is indistinguishable from pure noise. Then you train a network to reverse that process, step by step, learning to denoise. And the deep reason Gaussian noise is the choice is exactly the convenience engine we just described. The noising process has clean closed-form properties, you can jump to any noise level analytically, the training objective is a simple squared-error loss. It's the same tractability that made least-squares the default for two centuries.



Corn

The image is melted into static and reassembled, and the static has a brand.



Herman

The static is Gaussian, by decree. Now, is that a problem? Here's the live fight. If your data itself is heavy-tailed, which a lot of real image data is, you have outliers, rare regions, unusual compositions, then the intuitive move is to match the noise to the data. Use a heavy-tailed noise distribution, like a Student's t, instead of a Gaussian. People have built those, heavy-tailed diffusion models, and they do improve fidelity in the tails, the rare stuff comes out more faithfully.



Corn

You said fight, which means someone disagrees.



Herman

A paper this year by Cherkaoui and colleagues takes exactly that trend apart. Their point is subtle and I had to sit with it. When you swap Gaussian noise for heavy-tailed noise, you're not just changing the noise. You're changing the estimation problem the network has to solve. And their bounds show the heavy-tailed version makes the statistical estimation harder, with less favorable sampling-error bounds. Their conclusion, near verbatim, is that this calls into question a growing design trend in generative modeling. The intuitive fix, match the noise to the data, is not free. It may cost you more in estimation than it buys you in tail fidelity.



Corn

So the two tails argue with each other.



Herman

There's an even stranger result on top of that. Shen and Zhao proved that when the denoising network is conditioned on the right variables, the noise law becomes a nuisance. In the exact-denoiser limit, the generated distribution is invariant to it. They tested it with a designed eight-times volatility shock in the mixing process and the output envelope moved by under thirteen percent. Which suggests that at the limit of a perfect denoiser, what noise you injected matters less than everyone designing these systems assumes.



Corn

That's almost insulting to the entire research program. Pour whatever you like into the image, the network learned the data either way.



Herman

At the limit, with the right conditioning, yes. The qualifier is that real denoisers aren't at the limit, so the noise choice matters in practice, but how much and in which direction is unsettled. That's where the field is right now. There is no consensus paper you can cite.



Corn

Okay. So inside generative models, the Gaussian is the chosen ingredient, and the argument is about whether the ingredient was chosen for the recipe or for the cook. Now connect it to the thing Daniel actually heard us talk about, synthetic data and model collapse.



Herman

This is where I think it gets interesting, because the collapse literature describes the failure in explicitly distributional language. Model collapse is defined as repeated training eroding distributional tails and homogenizing outputs. Fluency survives, tails go first. The middle of the distribution, the common cases, stays fine, and the rare cases, the unusual compositions, the edge examples, the long tail of human expression, get sanded off generation by generation. You end up with a model that's confidently fluent about an ever-narrowing world.



Corn

So collapse is the Gaussian's thin tails, arriving from the other direction.



Herman

Structurally the same error, arrived at differently. You didn't assume thin tails. You generated them. The output distribution of a generative model compresses toward the bulk because that's where the probability mass, and therefore the training signal, lives. And here's the finding that really got me. There's an ICML paper this year on sample selection bias, and it shows that the safeguard people use against collapse can become the mechanism that causes it. If you're in a setting where data can't be pooled, healthcare consortia, proprietary financial institutions, and each silo verifies samples against its own



Herman (continued)

local reference distribution, then selection preferentially keeps the samples that match the local manifold and prunes globally relevant tail modes.



Corn

The filter is shaped like the silo.



Herman

Yes. Each institution's quality check quietly deletes exactly the rare examples that the global distribution needed. The safeguard against collapse precipitates collapse. They prove it accelerates, with diversity decaying as a power law.



Corn

Corn's field test question, the one I always ask about tools: what does the system do when nobody's watching. What does it do when nobody's checking the tails?



Herman

It flattens them. Nobody's dashboard shows tail erosion, because nobody's dashboard is looking at the tails. Every metric that's easy to compute, average loss, average fluency, average quality, is an L-two flavored, bulk-of-the-distribution metric. Which, hold on, that's Taleb's whole point about econometrics, that it's dominated by standard deviations and measures grounded in the Gaussian family. We'll get there.



Corn

Wait. Before finance, there's a piece of this I know you're holding back, because I've seen you get excited about it. The weight matrices thing. Where the story inverts.



Herman

This is my favorite part of the whole episode. So we've been treating heavy tails as the good thing the Gaussian can't give you, in data. Now look inside the trained network itself. There's a body of theory, heavy-tailed self-regularization, applying random matrix theory to



Herman (continued)

the weight matrices of deep networks. The finding is that training sculpts the spectra of those matrices into heavy-tailed shapes. And the kicker: a Gaussian-looking weight spectrum is a sign of an under-trained or badly regularized layer. The Gaussian is the spectrum of a network that hasn't done its work. Training is, in this picture, the process of pushing away from Gaussian.



Corn

The model spends its whole training escaping the distribution we spent the first half of the episode praising.



Herman

In the weights, yes. And it's not just a diagnostic curiosity anymore, people are building training methods on it. AlphaPruning reads those spectral shapes to decide how much to prune each layer, and it prunes LLaMA-7B to eighty percent sparsity. There's a modified optimizer, HTMuon, built on the observation that the standard Muon update suppresses heavy-tailed spectra, and the heavy-tail-corrected version cuts pretraining perplexity. And layerwise learning rates guided by heavy-tailedness moved a one-billion-parameter model's zero-shot accuracy from about forty-seven percent to about forty-nine percent.



Corn

Two points of accuracy for following the tail.



Herman

On a billion-parameter model, from adjusting learning rates per layer based on a spectral shape metric. That's real engineering value from a distributional observation.



Corn

So hold the whole picture. In the data, heavy tails are what you want and what gets lost. In the noise process, Gaussian is the convenient default and the argument is whether swapping it out helps. In the weights, heavy tails are what training itself creates, and cultivating them deliberately improves the model. The Gaussian is scaffolding. The heavy tail is the building.



Herman

And now let's take the same idea out of AI entirely, because the purest version of this story happened in finance, and it has a body count.



Corn

The Gaussian copula.



Herman

Two thousand. David Li, at JPMorgan Chase, publishes a paper on default correlation using a copula function approach. A copula is actually a lovely object, it's the mathematical device for taking a set of marginal distributions and gluing them together with a dependence structure. It separates the question of how each thing behaves on its own from the question of how things move together. Li's formula used a Gaussian copula to model how defaults correlate, and it was adopted, in the industry's words, because the math was simple by Wall Street standards. You could price the correlation between mortgages in a CDO without ever looking at historical default data. You pulled the correlation from credit default swap prices instead.



Corn

Which is a circularity, isn't it. The market prices the risk using the formula, and the formula reads the risk off the market prices.



Herman

It's worse than circular, because the swap prices came from a period when house prices only ever went up. So the correlation being fed in was calibrated on a benign world. Then the structure itself failed. The Gaussian copula has zero tail dependence, which means it assigns essentially no probability to many things going wrong at the same time. Under a Gaussian, joint extremes are so improbable they're not worth pricing. When house prices fell, the true correlations exploded, defaults cascaded together, and instruments modeled as safe were suddenly worthless. One journalist later called it the formula that killed Wall Street.



Corn

Let me make sure the blame lands correctly, because I've heard two stories. One is that the Gaussian copula is structurally guilty, thin tails by design. The other is that the copula was fine and the inputs were garbage.



Herman

Keep both. They're not mutually exclusive, and the honest account needs both. The structural defect is real: zero tail dependence in a model of correlated default is a defect for the purpose it was used for. But the inputs were also spectacularly wrong. S&P's CDO Evaluator in two thousand one assumed a thirty percent correlation for asset-backed securities. The actual average correlation for mortgage pools over the preceding years was zero point zero six. Essentially no correlation, where a third had been assumed. And then a drilldown in two thousand six, right before the fall, found real correlations near eighty percent when markets were stressed.



Corn

Thirty percent assumed, point-oh-six actual in calm times, eighty percent in the storm. The model was wrong at rest and wrong under load, in opposite directions.



Herman

That's the cleanest summary I've heard anyone put on it. Wrong at rest, wrong under load. So yes, garbage in, garbage out, but also, the machine was built to dispose of a specific kind of garbage gracefully, and tail dependence wasn't it.



Corn

And this is Taleb's terrain. He's been saying for decades that the entire toolkit is Gaussian-flavored.



Herman

Taleb and Goldstein, in the forecasting journal, make the case with data I still find staggering. Close to twenty million pieces of economic data, and the fat tails persist under aggregation. Which is the bit that kills the standard defense, the one where you say, fine, individual observations are wild, but averaging tames them, the central limit theorem will save us. No. In real economic data the wildness survives aggregation. And their most famous figure: in ten thousand US stock market observations, seventy-eight percent of the kurtosis came from a single observation. One day did most of the work of being



Herman (continued)

extreme.



Corn

Which is the number we opened with, now with its machinery visible. If your toolkit is standard deviations and squared-error measures, all of it grounded in the Gaussian family, and one observation carries most of the shape of your data, then your instruments aren't measuring the thing that matters.



Herman

Taleb's deeper argument, the one I find hardest to shake, is about the reassurance mathematicians offer. The textbook says: power-law distributions with finite variance are amenable to Gaussian methods, owing to that finite variance. Taleb's response, from his fat-tails work, is that outside the idealized model world, there's no consequential difference between finite and infinite variance models when the tails are that fat. The license you thought finite variance gave you doesn't do anything in practice.



Corn

The permit expires the day you need it.



Herman

The finite-variance license is exactly the kind of thing that's technically true and practically useless, and I say that as someone who spent a career explaining small probabilities to people who were too frightened to hear them. The number was never the problem. The shape the number was sitting in was the problem.



Corn

You know what strikes me, looking back across the whole episode? Every one of these failures, the copula, the collapse, the overconfident forecaster, has the same silhouette. The middle is fine. Reports come in, the average looks right, the quality scores hold, fluency survives. And the rare thing, the joint extreme, the tail mode, was priced at zero or near zero, and it turns out to be the whole story.



Hilbert

Forty minutes, Herman. You said the histogram can't be trusted, the middle of it looks fine, and then you moved on to finance.



Herman

I stand by the histogram part.



Hilbert

So do I, mostly. My complaint is narrower. People confuse bell-shaped with normal because they've only ever seen histograms drawn with too few bins. Bin width does the work. I spent a weekend once fitting a curve to the wait times at the post office on Jaffa Road. I was convinced it would say something about queueing theory. It did not. The waits were not normal, and it took me a full Sunday of binning to accept it.



Corn

What were they?



Hilbert

Log-normal. Products of little delays, each one multiplying the last, exactly as you two described, though you took the scenic route. I still have the histogram. It's in my wallet, folded into eighths, and it has been through the wash twice. I told the clerk at the time, a man named Amos, who could tell you the exact minute the morning line peaked and never once got it wrong by more than two minutes, that his waits were log-normal and not normal, and he gave me a stamp booklet to stop the lecture. I consider that a fair price on both sides. When I was doing quality control we settled arguments with less.



Herman

The fact that it's log-normal is actually perfect, that's the multiplicative queue, each delay scaling the next.



Hilbert

It's a good booklet. Six stamps.



Corn

You've been carrying a washed histogram of the Jaffa Road post office for years and your regret is the binning.



Hilbert

The binning was correct. The clerk was a decent man about it. That's more than most.



Corn

The one thing, then, if you take one thing from today. A bell-shaped middle tells you almost nothing, because half of statistics is bell-shaped in the middle. The Gaussian earns its place only when you know why it's there, a genuine sum of many small independent effects, or it's a deliberate convenience you're auditing. The tails are the whole question, and the tails are the part you can't see.



Herman

And the sharpening on that: the failure never announces itself in the middle. The CDO kept paying. The collapsed model stayed fluent. The warnings live exclusively in the part of the distribution you have to go looking for.



Corn

So the open question I'd leave with Daniel. The Gaussian is the scaffolding of modern AI, the noise, the losses, the defaults, and the heavy tails are the building. As synthetic data becomes a larger and larger share of what models learn from, the tail-erosion problem compounds on itself, and the invisible default is sitting underneath it, quietly assuming the rare thing away. What happens when the scaffolding gets mistaken for the structure, at scale?



Herman

That's the episode to watch for. Meanwhile, our thanks as always to Hilbert Flumingtop, producing this show from a desk that, we now know, contains a folded histogram and the remnants of a stamp booklet.



Corn

For more along these lines, there's episode twelve, The AI Breakthrough; episode fifteen, AI Gets Personal; and episode three, Safetensors or something else. This has been My Weird Prompts. If today's episode earned it, leave us a review, it's how the show finds curious people.



Herman

We'll be back soon. See you then.

SOURCES

What the research for this episode read before the script was written. Primary sources first.

1. **Wikipedia, Normal distribution, last edited 2026-10-29**
   [primary reference]
   https://en.wikipedia.org/wiki/Normal_distribution

2. **Wikipedia, Multivariate normal distribution, last edited 2026-11-02**
   https://en.wikipedia.org/wiki/Multivariate_normal_distribution

3. **Wikipedia, Log-normal distribution, last edited 2026-11-02**
   https://en.wikipedia.org/wiki/Log-normal_distribution

4. **Wakatsuki & Shimazaki, Self-Regulating Annealing in Heavy-Tailed Diffusion Models, arXiv:2606.01645v1, submitted 2026-06-01**
   https://arxiv.org/abs/2606.01645v1

5. **Cherkaoui, Halconruy & Ocello, Do Heavy Tails Help Diffusion? On the Subtle Trade-off Between Initialization and Training, arXiv:2605.13175v2, submitted 2026-05-13, revised 2026-06-01**
   https://arxiv.org/abs/2605.13175v2

6. **Weng & Liu, SHADE: Shrinkage via Halted Alpha-stable Denoising for Heavy-Tailed Probabilistic Time Series Forecasting, arXiv:2610.26358v1, submitted 2026-10-19**
   https://arxiv.org/abs/2610.26358v1

7. **Shen & Zhao, Denoising Subordinated Probabilistic Models, arXiv:2607.19218v1, submitted 2026-07-21**
   https://arxiv.org/abs/2607.19218v1

8. **Baule, Generative modelling with jump-diffusions, arXiv:2503.06558v2, submitted 2025-03-09, revised 2026-02-06**
https://arxiv.org/abs/2503.06558v2

9. **Qiao, Du, Liu et al., When Sample Selection Bias Precipitates Model Collapse, arXiv:2606.13732v2, submitted 2026-06-11, revised 2026-07-02, ICML 2026**
https://arxiv.org/abs/2606.13732v2

10. **Myakala, Namburi, Keragodu Jayaramu et al., SynthSentry: Detecting Synthetic Data Contamination in Language Model Training Data, arXiv:2609.12353v1, submitted 2026-09-11**
https://arxiv.org/abs/2609.12353v1

11. **Mahoney, Traditional and Heavy Tailed Self Regularization in Neural Network Models, PMLR v97, 2019**
https://proceedings.mlr.press/v97/mahoney19a.html

12. **Mahoney, Implicit Self-Regularization in Deep Neural Networks, JMLR 22 (2021), submitted 4/20, published 6/21**
https://www.stat.berkeley.edu/~mmahoney/pubs/htsr_20-410.pdf

13. **AlphaPruning: Using Heavy-Tailed Self Regularization Theory for Improved Layer-wise Pruning of Large Language Models, NeurIPS 2024**
https://papers.nips.cc/paper_files/paper/2024/hash/10fc83943b4540a9524af6fc67a23fef-Abstract-Conference.html

14. **HTMuon: Improving Muon via Heavy-Tailed Spectral Correction, ACL Findings 2026**
https://aclanthology.org/2026.findings-acl.1819/

15. **He et al., One LR Doesn't Fit All: Heavy-Tail Guided Layerwise Learning Rates for LLMs, PMLR v306**
https://proceedings.mlr.press/v306/he26ak.html

16. **Royal Statistical Society, *Significance*, The formula that killed Wall Street, 2012**
https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2012.00538.x

17. **Andy Jones, Copulas and the Global Financial Crisis**
https://andrewcharlesjones.github.io/journal/copulas-and-gfc.html

18. **Samuel Watts, The Gaussian Copula and the Financial Crisis**
https://samueldwatts.com/wp-content/uploads/2016/08/Watts-Gaussian-Copula_Financial_Crisis.pdf

19. **Cifuentes & Katsaros, The One-Factor Gaussian Copula Applied To CDOs: Just Say NO**
https://www.dii.uchile.cl/wp-content/uploads/2011/05/Cifuentes_Katsaros.pdf

20. **Taleb & Goldstein, The problem is beyond psychology: The real world is more random than regression analyses, *International Journal of Forecasting*, 2012**
https://fooledbyrandomness.com/Taleb-Goldstein-IJF2012.pdf

21. **Taleb, Statistical Consequences of Fat Tails, arXiv:2001.10488v2**
https://arxiv.org/abs/2001.10488v2

22. **Taleb, Option traders use (very) sophisticated heuristics, never the Black**
Scholes, Merton formula, *JEBO*
https://www.sciencedirect.com/science/article/abs/pii/S0167268110001927

23. **Taleb, *Complexity*, June 2008**
https://fooledbyrandomness.com/complexityAugust-06.pdf