Chapter 11 - Probability calibration under Fat Tails

Marcos Costa Santos Carreira

Technical Incerto Reading Club - 29-Jul-2021

Contents

One-liners

- "In nature there are neither rewards nor punishments—there are consequences." Robert G. Ingersoll, Some Reasons Why (1881)
- Ludic fallacy
- Distributional uncertainty
- Tail clipping

Magicians And Mathematicians

- https://wilmott.com/magicians-and-mathematicians/
- What is the probability of the Ace of Hearts?
- Are those the only two possible answers? Either one in 52 or 100%?

Domains

- Restricted domains and rewards in line with probability forecasts are mostly artificial constructs (classification problems are rarely simple)
- And they are often misused
- https://en.wikipedia.org/wiki/Fairness_(machine_learning)
- In real life outliers are people
- https://www.laserfiche.com/ecmblog/the-sad-story-of-mr-null-and-little-bobby-tables/

More about domains

- What is an efficient encoding of the decision domain?
- Thresholding ("There be dragons") and branching only when needed
- Assumes local knowledge (no global knowledge)
- Oversampling ("The Flood") - adjust representation of imbalanced classes
- In real life decisions accrue over time
- In real life the collective (species) benefits/suffers from individual decisions in a non-linear way

Impact

$$\mathrm{I}_1(f, g, K) = \int_K^\infty f(x) g(x)\, dx$$

$$\mathrm{I}_2(f, g, K) = g(K) \int_K^\infty f(x)\, dx$$

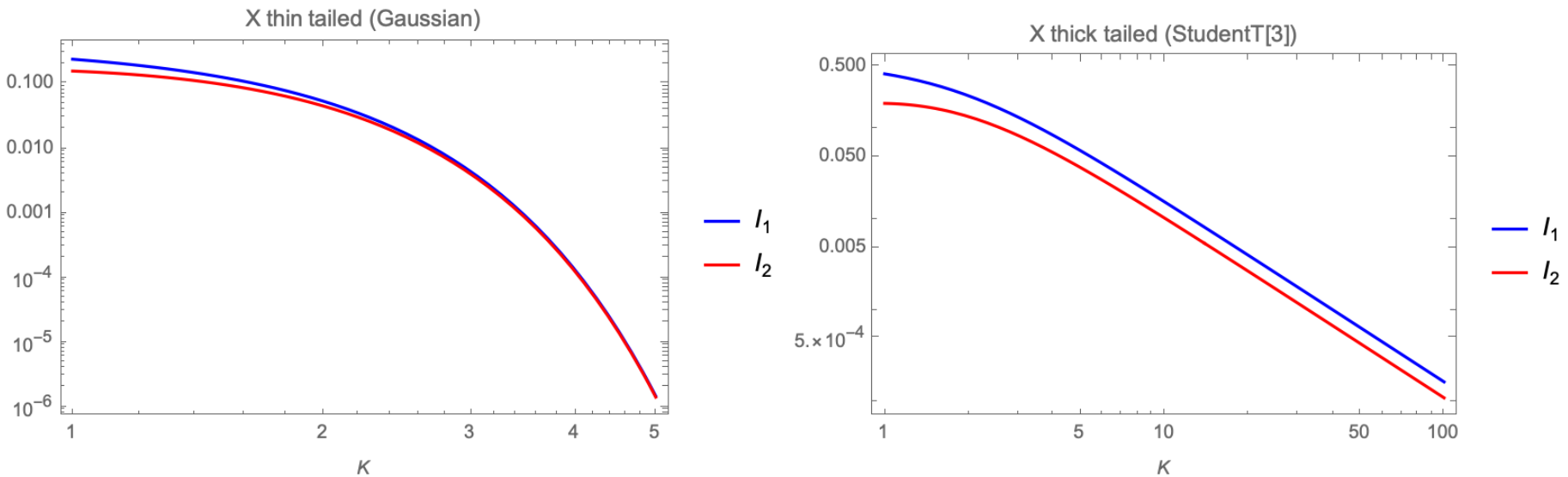


Figure: Impacts for $g(x) = x$

Tail survival

$$P(X > K) = +\frac{1}{2} - \frac{\sqrt{3}K}{\pi\left(K^2+3\right)} - \frac{\tan^{-1}\left(\frac{K}{\sqrt{3}}\right)}{\pi}$$

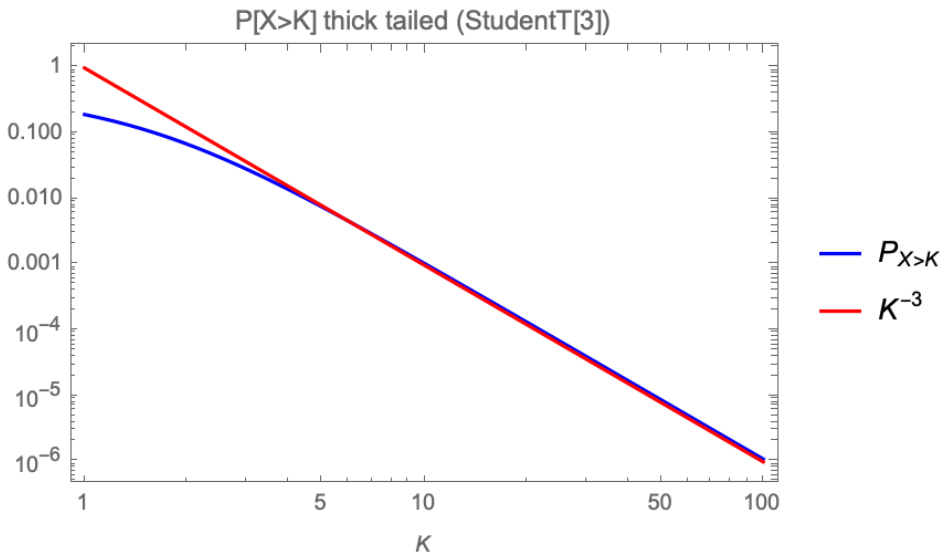


Figure: Tail Survival

LogNormal Bifurcation

- LogNormal:
  $\mathrm{fl}(x, \sigma, K) := \mathrm{PDF}\left[\mathrm{LogNormalDistribution}\left[\log(X_0) - \frac{\sigma^2}{2}, \sigma\right], x\right]$:
  Mean X_0 and Variance $X_0^2\left(\exp\left(\sigma^2\right) - 1\right)$
- Call Option: $\mathbb{E}(X > X_0) = \frac{1}{2} X_0 \left(1 + Erf\left(\frac{\sigma}{2\sqrt{2}}\right)\right)$
- Binary Option: $P(X > X_0) = \frac{1}{2}\left(1 - Erf\left(\frac{\sigma}{2\sqrt{2}}\right)\right)$

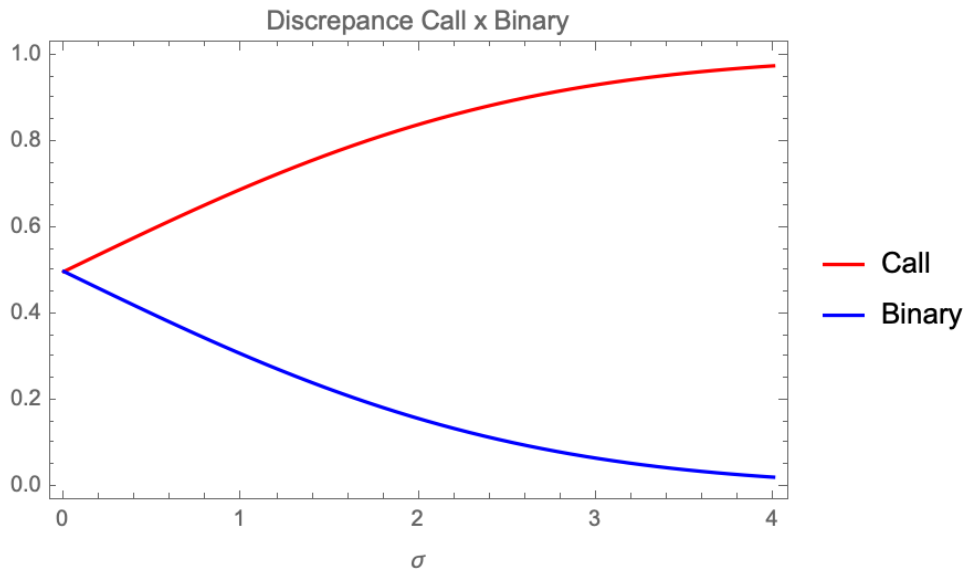


Figure: LogNormal bifurcation

Ergodicity Economics

- Ole Peters `https://ergodicityeconomics.com`
- Coin Game with Multiplicative Dynamics
- Log[Wealth] on horizontal axis, n (throws) from back to front, probability on z axis
- Expected Value (in Red) goes up but mass/mode of probability is on losses

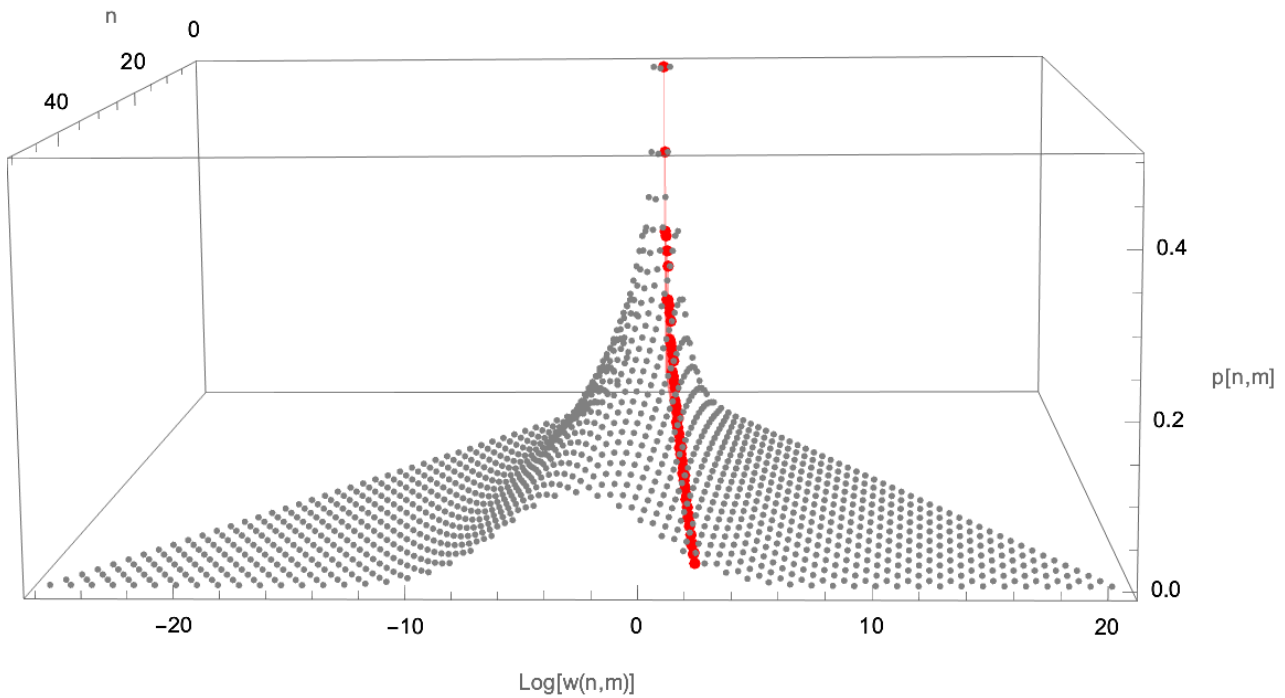


Figure: Coin Game

Distributional Uncertainty

- $I_1 = \frac{\nu^{\nu/2}\left(K^2+\nu\right)^{\frac{1}{2}-\frac{\nu}{2}}}{(\nu-1)B\left(\frac{\nu}{2},\frac{1}{2}\right)}$
- $I_2 = K\left(\frac{1}{2} - \frac{K\Gamma\left(\frac{\nu+1}{2}\right){}_2F_1\left(\frac{1}{2},\frac{\nu+1}{2};\frac{3}{2};-\frac{K^2}{\nu}\right)}{\sqrt{\pi}\sqrt{\nu}\Gamma\left(\frac{\nu}{2}\right)}\right)$

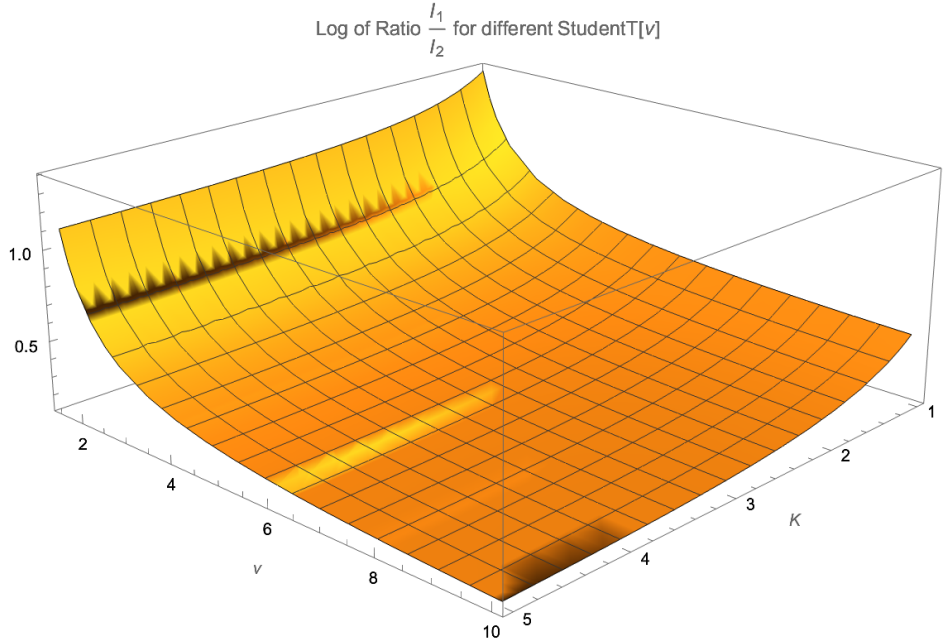


Figure: Ratio as a function of ν

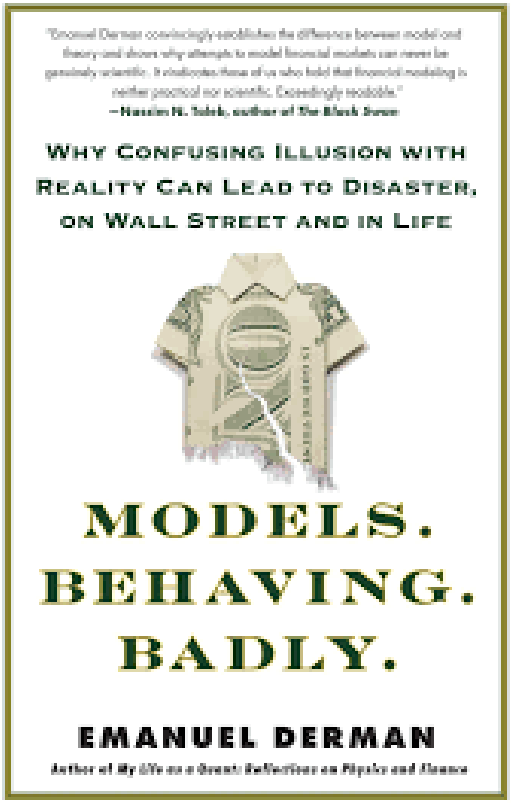
Rules and rulers

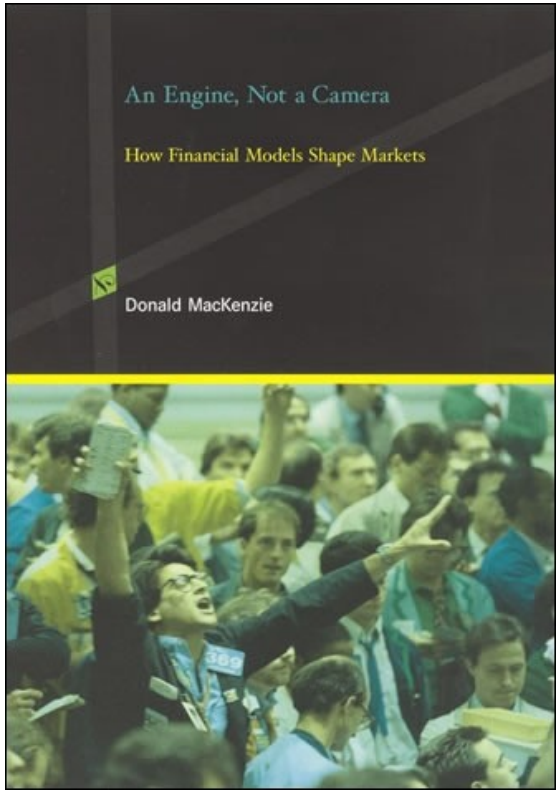
Metric	Notes
Cumulative P&L	Real world, survival; but ... counterfactuals? FBR?
Tally of Bets	No time structure, dependent on definition of ranges
Brier Score	Is not very adequate for very rare (frequent) events
M4 Score	sMAPE (good)
M5 Score	RMSSE (better) - scaled errors

- Cumulative P&L: Absorbing barrier (stop loss, increase in risk limit consumption)
- Brier Score: Decomposition in Calibration + Refinement
- M5: Link to M5 Accuracy Competition paper
- Not all the math in the chapter was reproduced. To Do: 11.7.4 (not fluent in the Kummer confluent hypergeometric function)

LTCM

- Tally of bets led to false predictions
- Finance - beware (models will change the world):





Tail Clipping

- Zeynep, Apr 2020:
  - "The most important function of epidemiological models is as a simulation, a way to see our potential futures ahead of time, and how that interacts with the choices we make today."
  - "Instead, we need to focus on the branches representing the worst outcomes, and prune them with all our might."
- NNT, Commentary 11.7:
  - Risk management is about changing the payoff function rather than making good forecasts.

Epidemiology

- We do not need to be precisely right 1 month from now - we just need to know whether the slope will increase
- We measure success not by nailing the death count, but by making every death count
- Individual decisions change the the probabilities and the payoff (no quick resolution, hospitals full, etc.) to others

What this talk was about anyway?

Risk Management is not Prediction

- Probabilities do not map into consequences and it's hard to extrapolate success on forecasting to rare events
- We survive not by improving the forecasting of frequent events, but by clipping tails

By Muhammad Mahdi Karim - Own work, GFDL 1.2, https://commons.wikimedia.org/w/index.php?curid=9543449