Bayesian Modeling and PyMC

NCSU IAA 2025

Evan Wimpey
PyMC Labs

AGENDA

01 **Introduction**
- Evan
- PyMC (the OSS)
- PyMC Labs (the company)

02 **Bayes**
- Refresher on Bayesian concepts
- Why Bayes matters to companies
- Why it matters to you!

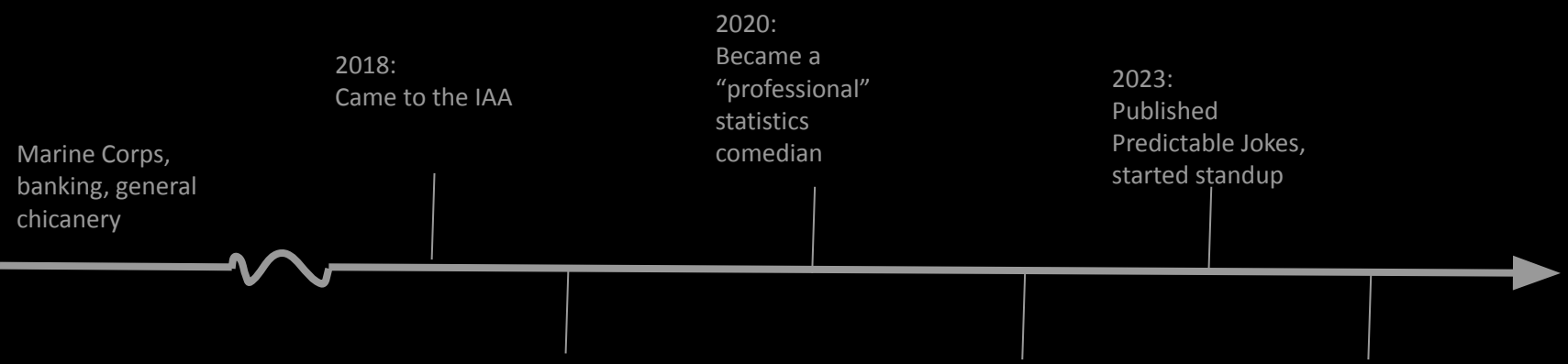
03 **PyMC Basics**
- Getting Started
- Modeling
- Other Modeling
- More Modeling

04 **PyMC Advanceds**
- Examples
- Documentation
- Community

05 **CLOSING**
- Questions
- Jokes
- Thank You

01

Introduction


Marine Corps, banking, general chicanery
2018: Came to the IAA
2019: Started as Data Scientist at Elder Research
2020: Became a “professional” statistics comedian
2021: Promoted to Director
2023: Published Predictable Jokes, started standup
2024: Started at PyMC Labs

PyMC is a probabilistic programming library for Python that allows users to build Bayesian models with a simple Python API and fit them using Markov chain Monte Carlo (MCMC) methods.


PyMC
Labs
THE BAYESIAN CONSULTANCY
www.pymc-labs.com

02

Bayes

A refresher?

What do we do here in “data science”

Throw stuff at an LLM?

Throw data into XGBoost?

What do we do here in “data science”

~~Throw stuff at an LLM?~~

~~Throw data into XGBoost?~~

Statistical Analysis!

1. Parameters
2. Data
3. Models

Frequentist Statistical Analysis

Data are random.

Parameters are fixed (even though they are unknown)

$P(y|\theta)$

Probability of our data conditional on some unknown parameters.

Bayesian Statistical Analysis

Data are fixed (once we collect them).

Parameters are random.

$P(\theta|y)$

Probability of parameters given some data (like we think!)

Bayes Formula

Data Likelihood

Prior Probability

$$P(\theta|y) = \frac{P(y|\theta) \times P(\theta)}{P(y)}$$

Posterior Probability

Normalizing Constant (the thing that makes this a pain in the

So Why Use Bayesian Analysis?

Quantifying Uncertainty	• For each parameter, or jointly across parameters, know how "sure" the model is • Smarter, risk based decisions
Modular Model Building	• Must explicitly formulate a model • Easier to explain and communicate • Can add and subtract and transform and anything!
Incorporating Prior Knowledge	• Most orgs know *something* that isn't in the data
Small data, big data, all good	• Works well with any data size • No "statistical significance" arbitrary tests, just wider estimates

Who uses Bayes?

- Marketing
- Risk (insurance, banking)
- Hard Sciences
- Economic modeling
- Image classification
- Hard math problems (like ODEs)
- Manufacturing optimization
- Maritime and weather
- Securities pricing
- Sports analytics
- Cybersecurity
- Everything, basically

Evan's Intuition

- Lost in Geneva
- Flipping a Coin
- Medical Testing

03

PyMC