Scene Guide to Information Geometry Visualization

Math-To-Manim

January 23, 2025

Abstract

This guide explains the animated visualization of information geometry concepts, connecting 3D probability manifolds with fundamental statistical theory. Each section breaks down complex mathematical ideas through their visual representations.

Scene 1: Cosmic Introduction - The Probability Universe

What You See:

The animation opens in a vast **3D starfield** with 400 stars in 8 depth layers, creating a mesmerizing parallax effect. A glowing title appears: "Information Geometry: The Landscape of Probability" in purple-to-teal gradient. As the title shrinks to the upper corner, the camera begins slowly orbiting through space, revealing different perspectives of the starfield.

Concept Explanation:

Imagine each star as a unique **probability distribution** in a cosmic space of possibilities. The depth layers represent:

- **Foreground:** Common distributions (Normal, Binomial)
- **Midground:** Complex statistical models
- **Background:** Abstract probability measures

The parallax effect mimics how statisticians view relationships between distributions - nearby distributions appear distinct while distant ones blur together. The camera's orbit represents our changing perspective when analyzing statistical relationships.

Key Takeaway:

Just as stars exist in cosmic structures, probability distributions live in a mathematical space called a **statistical manifold**. The animation positions us as cosmic explorers navigating this probability universe.

Scene 2: Probability Manifolds - Mountains of Likelihood

What You See:

Two glowing 3D surfaces emerge from the starfield:

- **Blue Mountain ():** Gentle bell curve centered at (0,0)
- **Red Volcano ():** Sharper peak shifted to (1,0.5)

The camera circles these structures, revealing their terrain. Surface transparency lets stars shine through, maintaining cosmic context.

Concept Explanation:

These surfaces represent **probability density functions** over a 2D parameter space:

$$(u, v) = e^{-(u^2+v^2)} \quad \text{(Gaussian-like)}$$
$$(u, v) = e^{-((u-1)^2+(v-0.5)^2)} \quad \text{(Offset peak)}$$

- **X/Y Coordinates:** Parameters controlling distribution shape
- **Height (Z):** Probability density at those parameters
- **Color:** Distinguishes different distribution families

The camera orbit demonstrates how statistical manifolds exist in **curved spaces** - ordinary "flat" geometry doesn't apply here.

Key Takeaway:

Probability distributions aren't just equations - they're **geometric landscapes** we can navigate. The peaks and valleys represent areas of high/low probability density.

Scene 3: KL Divergence - The Probability Compass

What You See:

50 golden particles appear on the blue -surface, then:

- Each particle displays a red/blue value label
- Equation appears: $\mathrm{KL}(\mu \parallel \nu) = \mathbb{E}_\mu[\log \frac{d\mu}{d\nu}]$
- Particles turn red (ratio $>$1) or blue (ratio $<$1)

Concept Explanation:

The Kullback-Leibler (KL) divergence measures how different is from :

- **Golden Particles:** Random samples from
- **Color Coding:**
  - Red: $>$ at this location (overestimates reality)
  - Blue: $<$ at this location (underestimates reality)
- **Label Values:** $\frac{(x)}{(x)}$ ratio at sample points

KL divergence averages these ratios across all samples - red dominance means and are very different.

Key Takeaway:

KL divergence acts as a "probability compass" showing directions where our model () diverges from reality (). More red particles mean poorer model fit.

Scene 4: Logarithmic Lens - Flattening the Cosmos

What You See:

White 3D arrows shoot from particles to new positions:

- Arrows curve through space
- Particles warp vertically
- Surface geometry appears to flatten

Concept Explanation:

The logarithmic transform converts multiplicative relationships to additive:

$$\log \frac{(x)}{(x)} = \log(x) - \log(x)$$

- **Arrows:** Visualize log-transform mapping
- **Warping:** Complex curvature becomes simple slopes
- **Flattening:** Exponential family distributions become "flat" in log-space

This transformation reveals the **duality** between:

- Original space (nonlinear, curved)
- Log-space (linear, flat)

Key Takeaway:

Like cosmic glasses, the logarithmic lens simplifies complex probability relationships into linear geometry we can easily navigate.

Scene 5: Fisher Information - The Cosmic Fabric

What You See:

The blue -surface transforms:

- Surface develops red/blue color gradient
- Equation appears: $\mathcal{I}(\theta) = \mathbb{E}[(\partial_\theta \log p_\theta)^2]$
- Surface ripples with curvature changes

Concept Explanation:

Fisher information measures how "sensitive" a distribution is to parameter changes:

- **Red Regions:** High sensitivity (steep slopes)
- **Blue Regions:** Low sensitivity (gentle slopes)
- **Curvature:** Represents local information content

This creates a **cosmic fabric** where:

- Tight curves = High information
- Flat regions = Low information

Key Takeaway:

Fisher information weaves the "fabric" of our statistical universe - it defines how we measure distances and angles between probability distributions.

Scene 6: Sufficient Statistics - Data Alchemy

What You See:

- 100 white dots (raw data) form a grid
- Data transforms into 10 green triangles
- New arrangement preserves essential patterns

Concept Explanation:

Sufficient statistics extract key information:

- **White Dots:** Individual data points
- **Green Triangles:** Statistical summaries (means, variances)
- **Transformation:** Lossless compression of information

The grid-to-triangles shift demonstrates the **Data Processing Theorem**:

$$\text{Information} \ \geq \ \text{Statistics} \ \geq \ \text{Knowledge}$$

Key Takeaway:

Just as alchemists sought essence, sufficient statistics distill data to its informative core without loss - the ultimate data compression.

Scene 7: Final Synthesis - Cosmic Equations

What You See:

A grand equation appears:

$$\mathrm{KL}(\mu \parallel \nu) = \underbrace{\mathbb{E}_\mu[\log d\mu]}_{\text{Entropy}} - \underbrace{\mathbb{E}_\mu[\log d\nu]}_{\text{Cross-Entropy}}$$

The camera pulls back to reveal all elements united - starfield, surfaces, particles, and equations.

Concept Explanation:

The finale unites key concepts:

- **Entropy (Blue):** Disorder in
- **Cross-Entropy (Red):** Surprise when encounters
- **KL Divergence:** Difference between them

This mirrors the cosmic balance:

$$\text{Information Gain} = \text{Self-Knowledge} - \text{Reality Check}$$

Key Takeaway:

Information geometry reveals deep cosmic truths - even probability distributions must balance self-knowledge with reality.