Maximum Likelihood on the Sphere

Statistics 98

Thomas Kaminsky

tkaminsky@college.harvard.edu
Harvard University

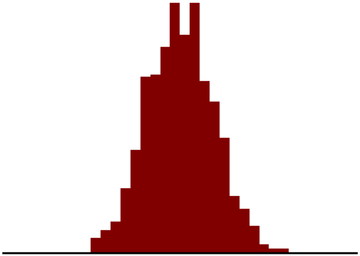
February 9, 2023

What is Maximum Likelihood Estimation?

What is Maximum Likelihood Estimation?

The Problem

- We often have some data that we believe follows a known distribution.
- We want to estimate the parameters of this distribution using our data.



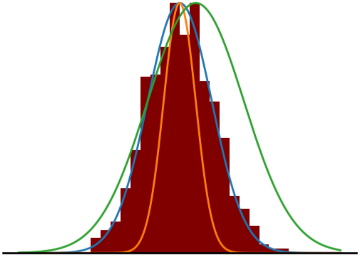
What is Maximum Likelihood Estimation?

The Problem

- We often have some data that we believe follows a known distribution.
- We want to estimate the parameters of this distribution using our data.

One Solution

- Construct the **likelihood**—the joint probability of all of our data given a choice of parameters.
- The maximum of this function is called the **Maximum Likelihood Estimate**, or **MLE**.



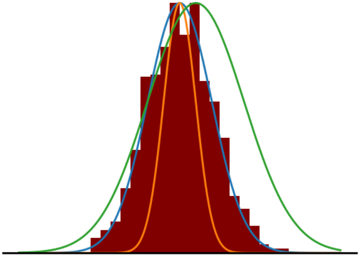
What is Maximum Likelihood Estimation?

The Problem

- We often have some data that we believe follows a known distribution.
- We want to estimate the parameters of this distribution using our data.

One Solution

- Construct the **likelihood**—the joint probability of all of our data given a choice of parameters.
- The maximum of this function is called the **Maximum Likelihood Estimate**, or **MLE**.



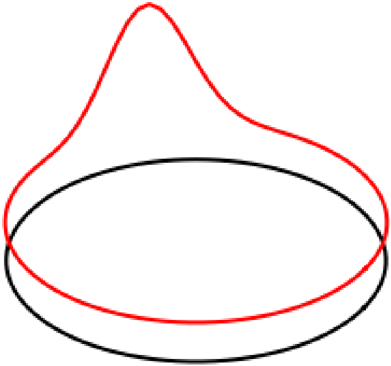
However, we have only worked with distributions whose support is Euclidean space. What if our data lies on a more complex manifold?

Directional Statistics

- Considers distributions whose support is over a **hypersphere**.
  - Angles or orientations.
  - Vector-valued quantities.

Directional Statistics

- Considers distributions whose support is over a **hypersphere**.
  - Angles or orientations.
  - Vector-valued quantities.
- The anologue of the Gaussian—the **von Mises-Fisher (vMF) Distribution**.



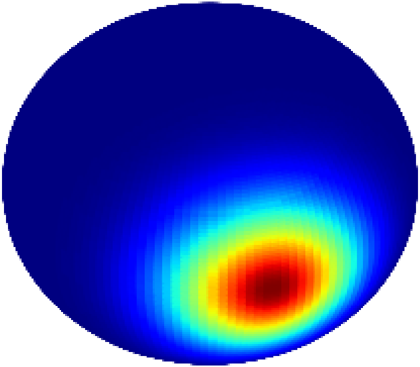
Directional Statistics

- Considers distributions whose support is over a **hypersphere**.
    - Angles or orientations.
    - Vector-valued quantities.
- The anologue of the Gaussian—the **von Mises-Fisher (vMF) Distribution**.

PDF given by

$$P(\mathbf{x}|\boldsymbol{\mu}) = N(\kappa)\exp\left(\kappa\boldsymbol{\mu}^T\mathbf{x}\right),$$

- Parameterized by *mean* unit vector μ and *concentration* $\kappa \geq 0$.
- Normalization constant $N(\kappa)$ found via numeric integration techniques.



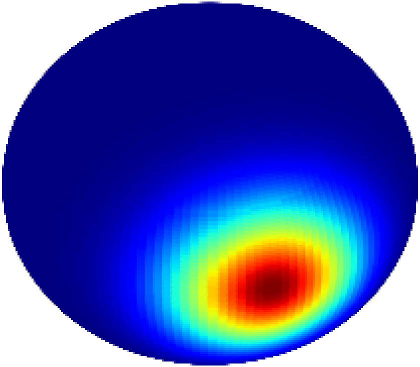
Directional Statistics

- Considers distributions whose support is over a **hypersphere**.
  - Angles or orientations.
  - Vector-valued quantities.
- The anologue of the Gaussian—the **von Mises-Fisher (vMF) Distribution**.

PDF given by

$$P(\mathbf{x}|\boldsymbol{\mu}) = N(\kappa)\exp\left(\kappa\boldsymbol{\mu}^T\mathbf{x}\right),$$

- Parameterized by *mean* unit vector μ and *concentration* $\kappa \geq 0$.
- Normalization constant $N(\kappa)$ found via numeric integration techniques.



The log-likelihood for some $\{\mathbf{x_i}\}_{i=1}^n$, dropping additive constants, is given by

$$\ell(\boldsymbol{\mu}|\mathbf{x}) \approx \kappa\boldsymbol{\mu}^T\left(\sum_{i=1}^n \mathbf{x_i}\right) = n\kappa\boldsymbol{\mu}^T\overline{\mathbf{x}}.$$

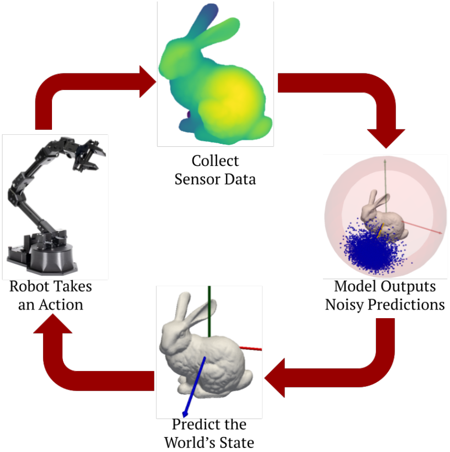
One Application: Robotic Vision

Who Cares?

One Application: Robotic Vision

Who Cares?

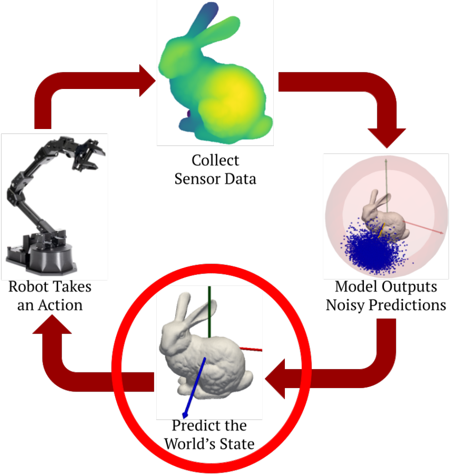
- Often in robotics, we see a pipeline such as the following:



One Application: Robotic Vision

Who Cares?

- Often in robotics, we see a pipeline such as the following:
- We need to make a good 'guess' about the state of the world.
- This is where we use the **MLE**.

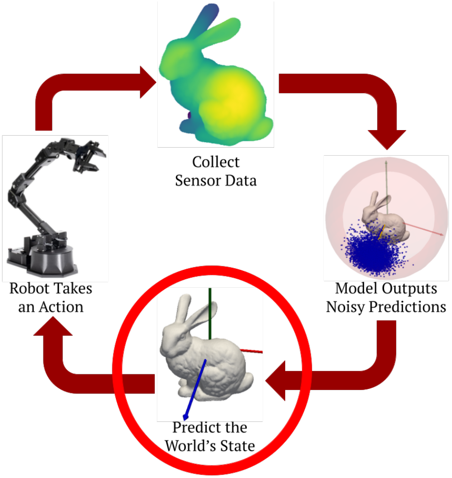


One Application: Robotic Vision

Who Cares?

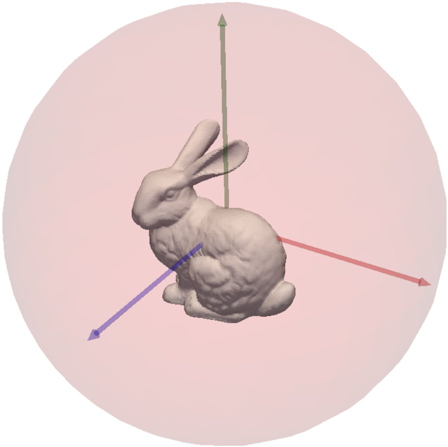
- Often in robotics, we see a pipeline such as the following:
- We need to make a good 'guess' about the state of the world.
- This is where we use the **MLE**.

Let's think about this process in a bit more detail.



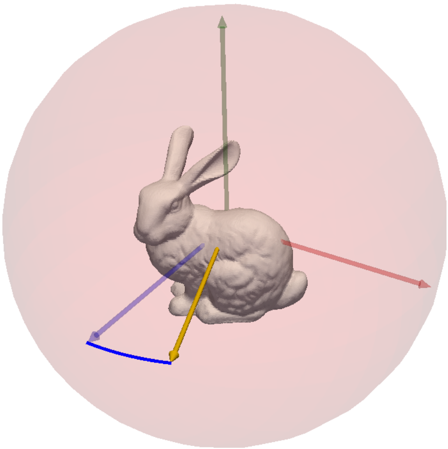
Our Problem: MLE on the Sphere

- We have a bunny. The robot knows how to work with it from some canonical orientation.



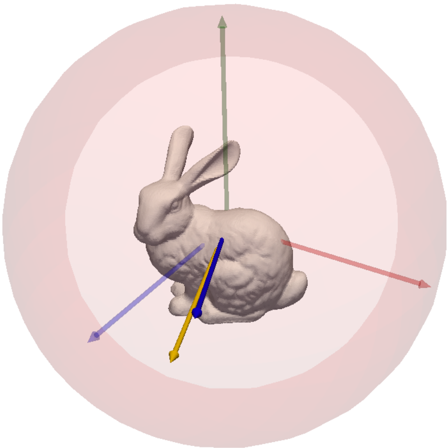
Our Problem: MLE on the Sphere

- We have a bunny. The robot knows how to work with it from some canonical orientation.
- But in the world, this orientation is different!
  - We want to find the rotated gold axis so we can match it with the blue axis in the robot's brain.



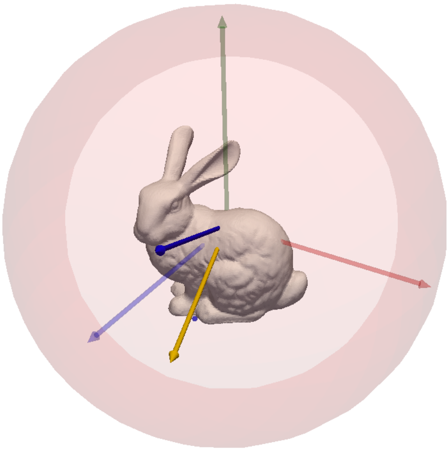
Our Problem: MLE on the Sphere

- We have a bunny. The robot knows how to work with it from some canonical orientation.
- But in the world, this orientation is different!
  - We want to find the rotated gold axis so we can match it with the blue axis in the robot's brain.
- You collect noisy sensor data:



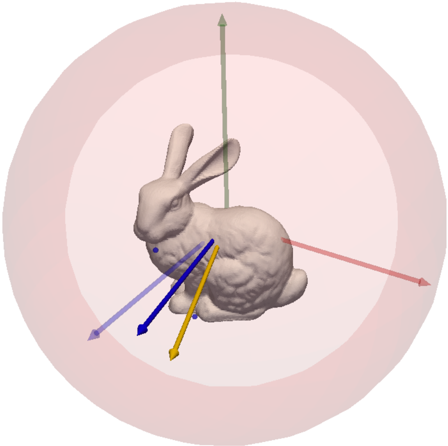
Our Problem: MLE on the Sphere

- We have a bunny. The robot knows how to work with it from some canonical orientation.
- But in the world, this orientation is different!
    - We want to find the rotated gold axis so we can match it with the blue axis in the robot's brain.
- You collect noisy sensor data:



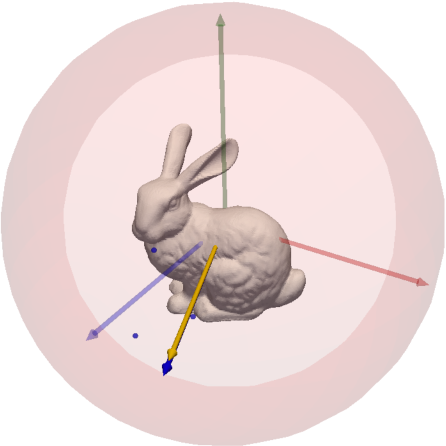
Our Problem: MLE on the Sphere

- We have a bunny. The robot knows how to work with it from some canonical orientation.
- But in the world, this orientation is different!
  - We want to find the rotated gold axis so we can match it with the blue axis in the robot's brain.
- You collect noisy sensor data:



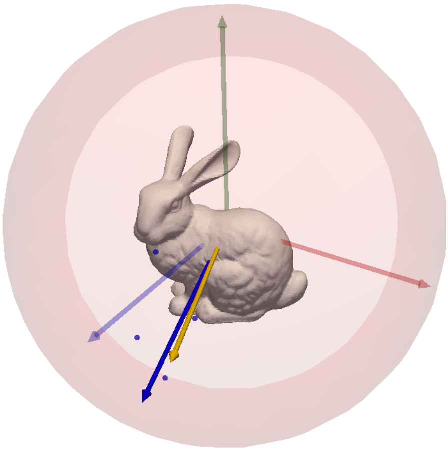
Our Problem: MLE on the Sphere

- We have a bunny. The robot knows how to work with it from some canonical orientation.
- But in the world, this orientation is different!
  - We want to find the rotated gold axis so we can match it with the blue axis in the robot's brain.
- You collect noisy sensor data:



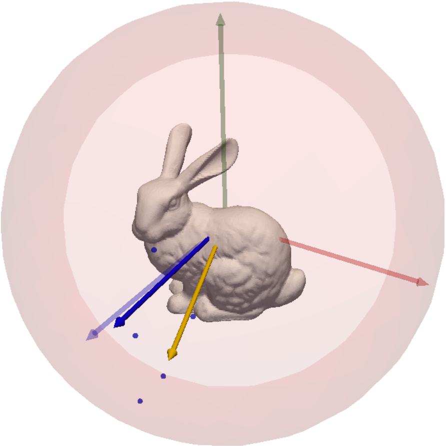
Our Problem: MLE on the Sphere

- We have a bunny. The robot knows how to work with it from some canonical orientation.
- But in the world, this orientation is different!
  - We want to find the rotated gold axis so we can match it with the blue axis in the robot's brain.
- You collect noisy sensor data:



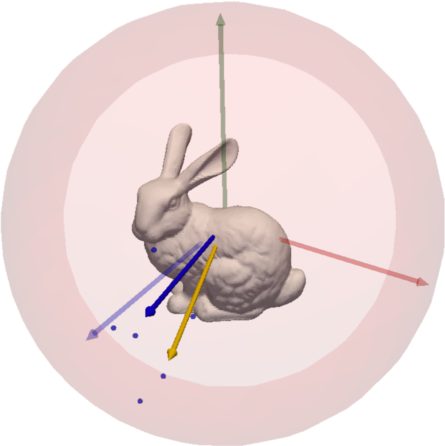
Our Problem: MLE on the Sphere

- We have a bunny. The robot knows how to work with it from some canonical orientation.
- But in the world, this orientation is different!
  - We want to find the rotated gold axis so we can match it with the blue axis in the robot's brain.
- You collect noisy sensor data:



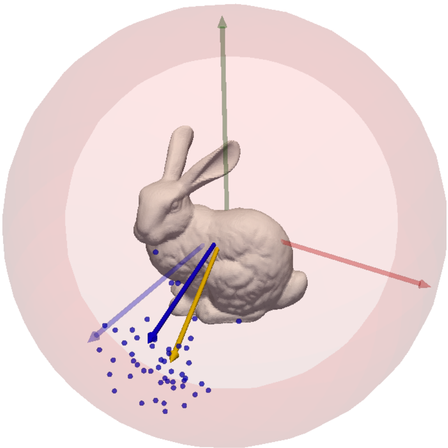
Our Problem: MLE on the Sphere

- We have a bunny. The robot knows how to work with it from some canonical orientation.
- But in the world, this orientation is different!
  - We want to find the rotated gold axis so we can match it with the blue axis in the robot's brain.
- You collect noisy sensor data:



Our Problem: MLE on the Sphere

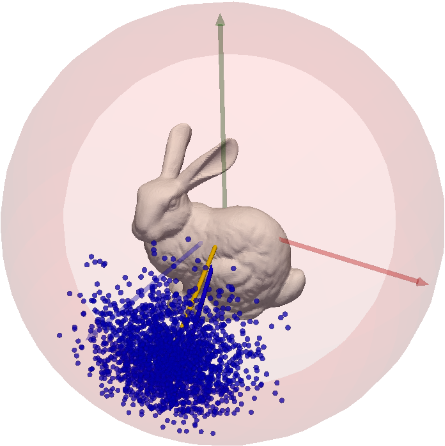
- We have a bunny. The robot knows how to work with it from some canonical orientation.
- But in the world, this orientation is different!
  - We want to find the rotated gold axis so we can match it with the blue axis in the robot's brain.
- You collect noisy sensor data:



Our Problem: MLE on the Sphere

- We have a bunny. The robot knows how to work with it from some canonical orientation.
- But in the world, this orientation is different!
  - We want to find the rotated gold axis so we can match it with the blue axis in the robot's brain.
- You collect noisy sensor data:
- Eventually, you have a large set of unit axis predictions.

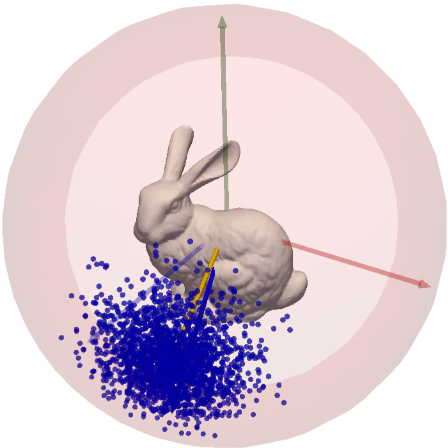
$$P(\boldsymbol{\mu}|\mathbf{x}) = N(\kappa)\exp\left(n\kappa\boldsymbol{\mu}^T\overline{\mathbf{x}}\right)$$



Our Problem: MLE on the Sphere

- We have a bunny. The robot knows how to work with it from some canonical orientation.
- But in the world, this orientation is different!
  - We want to find the rotated gold axis so we can match it with the blue axis in the robot's brain.
- You collect noisy sensor data:
- Eventually, you have a large set of unit axis predictions.
- We want to find a choice of μ that maximizes this data likelihood.
  - **Problem:** Can be made **arbitrarily large** by increasing $\mu^T\overline{x}$.

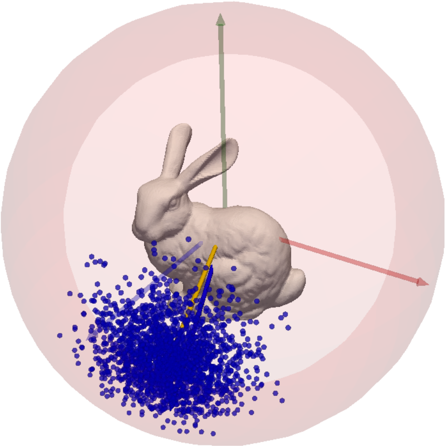
$$P(\boldsymbol{\mu}|\mathbf{x}) = N(\kappa)\exp\left(n\kappa\boldsymbol{\mu}^T\overline{\mathbf{x}}\right)$$



Our Problem: MLE on the Sphere

- We have a bunny. The robot knows how to work with it from some canonical orientation.
- But in the world, this orientation is different!
  - We want to find the rotated gold axis so we can match it with the blue axis in the robot's brain.
- You collect noisy sensor data:
- Eventually, you have a large set of unit axis predictions.
- We want to find a choice of μ that maximizes this data likelihood.
  - **Problem:** Can be made **arbitrarily large** by increasing $\mu^T \overline{x}$.
- We find that the problem is now a **constrained optimization**—maximize $P(\boldsymbol{\mu}|\mathbf{x})$ subject to $||\boldsymbol{\mu}|| = 1$.

$$P(\boldsymbol{\mu}|\mathbf{x}) = N(\kappa)\exp\left(n\kappa\boldsymbol{\mu}^T\overline{\mathbf{x}}\right)$$



Finding the MLE: Lagrange Multiplier Method

Recall from calculus that we can use Lagrange Multipliers to optimize a function subject to some constraint $g(\boldsymbol{\mu}) = 0$.

Finding the MLE: Lagrange Multiplier Method

Recall from calculus that we can use Lagrange Multipliers to optimize a function subject to some constraint $g(\boldsymbol{\mu}) = 0$.

Taking $g(\boldsymbol{\mu}) = ||\boldsymbol{\mu}|| - 1$, we find that our MLE is

$$\widehat{\boldsymbol{\mu}} = \pm \frac{\overline{\mathbf{x}}}{\|\overline{\mathbf{x}}\|}$$

This is simply the **mean** of our data, scaled to unit length.

Finding the MLE: Lagrange Multiplier Method

Recall from calculus that we can use Lagrange Multipliers to optimize a function subject to some constraint $g(\boldsymbol{\mu}) = 0$.

Taking $g(\boldsymbol{\mu}) = ||\boldsymbol{\mu}|| - 1$, we find that our MLE is

$$\widehat{\boldsymbol{\mu}} = \pm \frac{\overline{\mathbf{x}}}{\|\overline{\mathbf{x}}\|}$$

This is simply the **mean** of our data, scaled to unit length.

So, with a little calculus and basic statistics, we find that we can now tackle MLEs on complicated manifolds!

Finding the MLE: Lagrange Multiplier Method

Recall from calculus that we can use Lagrange Multipliers to optimize a function subject to some constraint $g(\boldsymbol{\mu}) = 0$.

Taking $g(\boldsymbol{\mu}) = ||\boldsymbol{\mu}|| - 1$, we find that our MLE is

$$\widehat{\boldsymbol{\mu}} = \pm \frac{\overline{\mathbf{x}}}{\|\overline{\mathbf{x}}\|}$$

This is simply the **mean** of our data, scaled to unit length.

So, with a little calculus and basic statistics, we find that we can now tackle MLEs on complicated manifolds!