Intern Final Presentation

Covariance Uncertainty: Estimation & Visualization

Aarush Khirwadkar

APL Summer Intern (FPS/KFA)

aarush.khirwadkar@jhuapl.edu

August 1, 2024

Agenda

About me

- Rising senior at UT Austin
- Studying Math & Economics, with minors in Business and Data Science, interested in:
  - Measure theory & Probability theory
  - Econometrics & Mathematical Statistics
  - Microeconomic & Game theory
  - Representation theory
  - Analytic & Existential Philosophy
- Previously: Wildfire Designs, texttobuy.xyz, Institute for Organizational Excellence, Innovations for Peace & Development
- Non-academic: Music, Table Tennis, Soccer

Problem Statement

Consider a bivariate normal distribution. KFA has legacy code for visualizing the 90% confidence region about the 1-σ ellipse.

Maximum Likelihood Estimation (MLE) is used to calculate the 1-σ ellipse and a Fisher Information approximation is used to estimate the 90% confidence region.

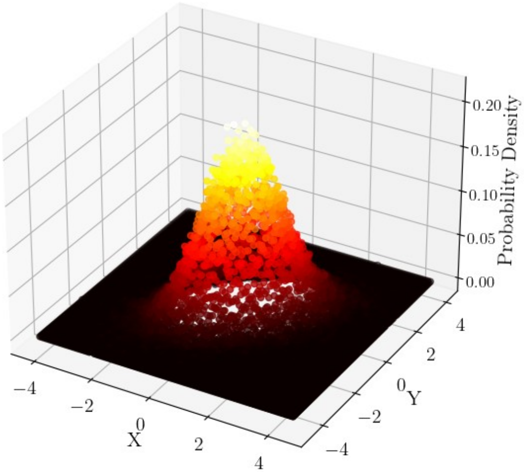
Challenge: Current code is slow, want to see if results are reproducible.

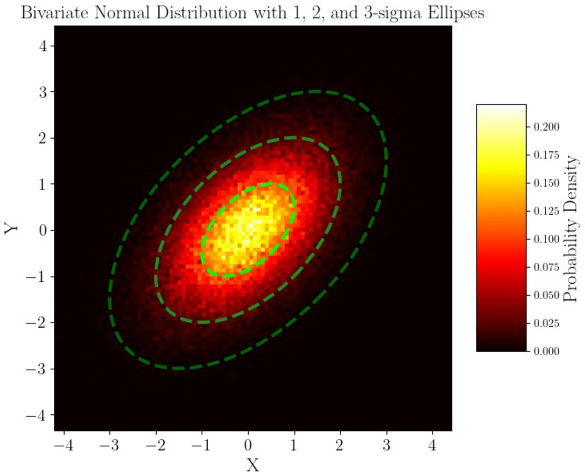
Two approaches to find the 90% region:

1. Using Fisher Information
2. Using Wishart distribution (ongoing)

Example of a bivariate normal distribution

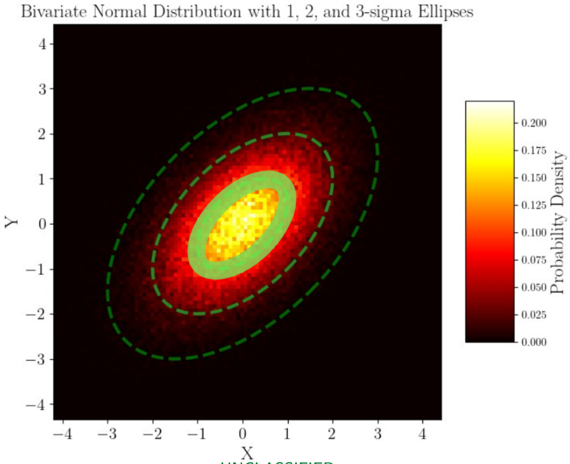
$$\text{mean}(x) = 0 = \text{mean}(y),\ \text{var}(x) = 1 = \text{var}(y),\ \text{cov}(x, y) = 0.5$$





Example of a 90% Confidence Region

We are interested in finding the region shaded in green, the **90% confidence region for the 1-σ ellipse**.



Overview of FIM Approach

In our assessment, for a bivariate normal distribution with independent and identically distributed (IID) samples, we:

- Estimated the MLE mean and covariances.
- Analytically calculate the Fisher Information Matrix (FIM).
- Calculate the asymptotic covariance matrix for the independent elements of the bivariate covariance matrix.
- Graph the 90% covariance ellipsoid and eliminate samples that violate the constraints on the bivariate covariance matrix.
- Graph the 90% confidence region about the 1-σ covariance ellipse.

Mathematical Setup

- Without loss of generality, set mean $= 0$ for both variables.
- Dataset:

$$\mathbf{x} := \begin{bmatrix} x_{1,a} & x_{1,b} \\ \vdots & \vdots \\ x_{n,a} & x_{n,b} \end{bmatrix},$$

  assumed to be drawn from a bivariate normal distribution, i.e., $\mathbf{x} \sim \mathcal{N}(0, \Sigma)$.
- MLE Covariance Matrix is assumed to be known and has values of:

$$\hat{\Sigma} = \begin{bmatrix} \hat{s}_a^2 & \hat{s}_{ab} \\ \hat{s}_{ab} & \hat{s}_b^2 \end{bmatrix},$$

  we are interested in estimating the uncertainty about this MLE covariance matrix.

Setup

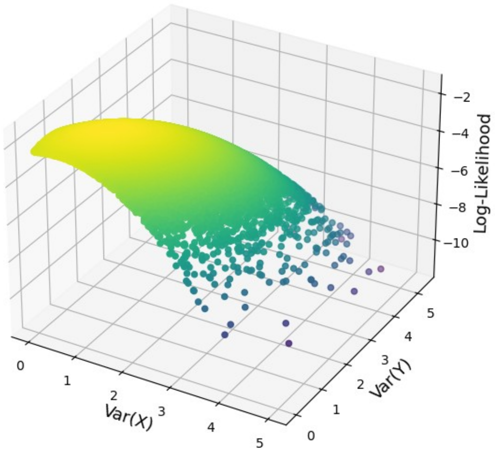
- Likelihood and log-likelihood functions:

$$L = (2\pi)^{-n} \cdot \left|\det(\hat{\Sigma})\right|^{-\frac{n}{2}} \cdot \exp\left(-\frac{1}{2}\sum_{j=1}^{n} \mathbf{x_j}\hat{\Sigma}^{-1}\mathbf{x_j}^{T}\right),$$

$$\ln(L) = -n\ln(2\pi) - \frac{n}{2}\ln(\hat{s}_a^2\hat{s}_b^2 - \hat{s}_{ab}^2) - \frac{1}{2}\frac{1}{(\hat{s}_a^2\hat{s}_b^2 - \hat{s}_{ab}^2)}\sum_{j=1}^{n}\left[\hat{s}_a^2(x_{j,b}^2) + \hat{s}_b^2(x_{j,a}^2) - \hat{s}_{ab}(2x_{j,a}x_{j,b})\right].$$

Example visualization of log-likelihood function

$$\text{var}(x) = 1,\ \text{var}(y) = 1,\ \text{cov}(x, y) = 0.5$$



Fisher Information Matrix (FIM)

The **Fisher Information Matrix** ($\mathbf{I}(\Theta)$) is the matrix containing entries

$$-\mathbb{E}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln L\right],$$

where θ_i, θ_j are the parameters (when we maximize log-likelihood). The **observed FIM** ($\mathbf{I}(\hat{\Theta}_{\mathrm{ML}})$) is the FIM evaluated with the observed data, without the expectation.

Cramér–Rao Lower Bound (CRLB)

CRLB describes the lowest variance for a biased estimator Θ. Namely,

$$\mathrm{Var}(\hat{\Theta}_{\mathrm{ML}}) \equiv \left(\frac{\partial}{\partial \Theta}\mathbb{E}(\hat{\Theta}_{\mathrm{ML}})\right)^2 \left[\mathbf{I}(\hat{\Theta}_{\mathrm{ML}})\right]^{-1},$$

where $\mathbf{I}$ is the Fisher Information of the parameter. When $\hat{\Theta}_{\mathrm{ML}} = \hat{\Sigma}_{\mathrm{ML}}$,

$$\mathrm{Var}(\hat{\Sigma}_{\mathrm{ML}}) \equiv \left(\frac{n-1}{n}\right)^2 \left[\mathbf{I}(\hat{\Sigma}_{\mathrm{ML}})\right]^{-1},$$

i.e., the inverse of the **observed FIM** is an estimator of the asymptotic covariance matrix.

Visualizing the Asymptotic Covariance in 3-d space

Using the asymptotic covariance matrix, we can apply the Cholesky decomposition to generate the 90% covariance ellipsoid. Namely,

1. Apply Cholesky decomposition:

$$\mathrm{Var}(\hat{\Theta}_{\mathrm{ML}}) \equiv \left(\frac{n-1}{n}\right)^2 \left[\mathbf{I}(\hat{\Theta}_{\mathrm{ML}})\right]^{-1} = \mathbf{L}\mathbf{L}^T,$$

   where $\mathbf{L}$ is a lower triangular matrix.

2. Randomly sample points on the surface of the unit sphere with center $(\hat{s}_a^2, \hat{s}_{ab}, \hat{s}_b^2)$. Arrange them into a matrix $\mathbf{P} \in M_{3\times\mathrm{numpoints}}(\mathbb{R})$.

3. Transform $\mathbf{P}$:

$$\mathbf{P} \mapsto \sqrt{\chi^2_{0.9,3}}\mathbf{L}\mathbf{P}.$$

Validity Criteria

We need:

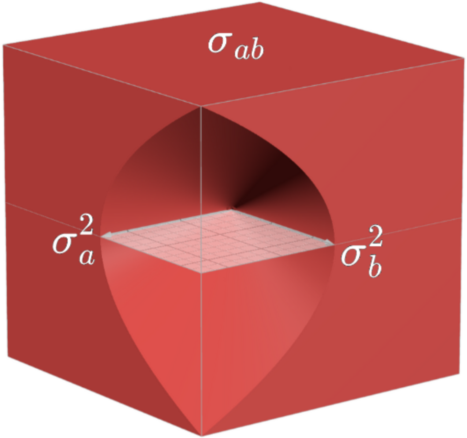
- Each transformed point to form a positive-definite matrix:

$$(\hat{s}_a^2, \hat{s}_{ab}, \hat{s}_b^2)^T \mapsto \begin{bmatrix} \hat{s}_a^2 & \hat{s}_{ab} \\ \hat{s}_{ab} & \hat{s}_b^2 \end{bmatrix}.$$

- $\hat{s}_a^2 > 0$, $\hat{s}_b^2 > 0$
- $|\rho_{ab}| < 1$, where $\rho_{ab} = \frac{\hat{s}_{ab}}{\sqrt{\hat{s}_a^2 \hat{s}_b^2}} \iff \hat{s}_a^2 \hat{s}_b^2 - \hat{s}_{ab}^2 > 0$

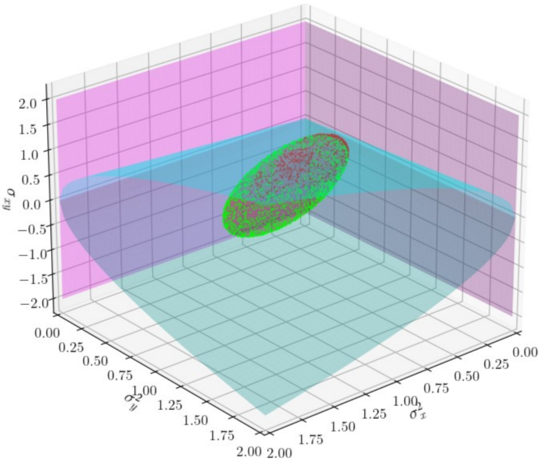
Valid Region

A sample is valid IFF it is in in the ‘carved out’ region. If it is in the ‘solid’ region, it is invalid.

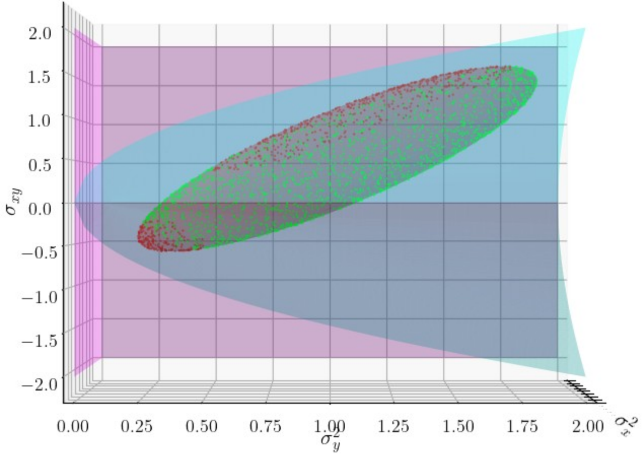


Visualization of 90% Covariance Ellipsoid

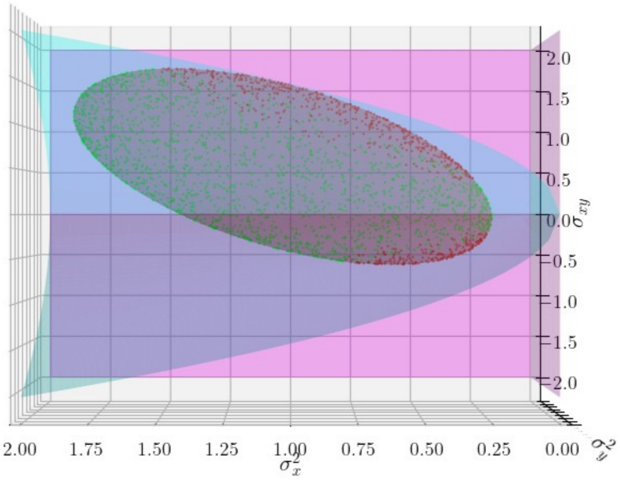
Valid samples in green, invalid samples in red.



Visualization of 90% Covariance Ellipsoid

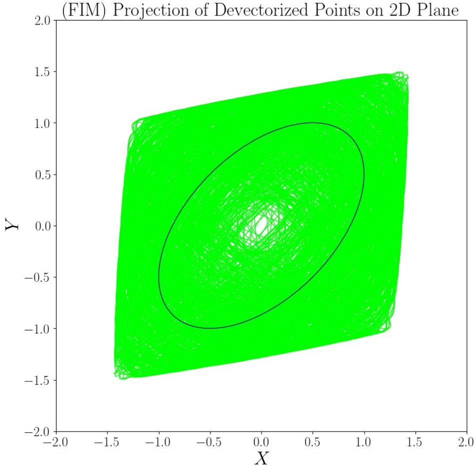


Visualization of 90% Covariance Ellipsoid



Plotting Ellipsoid on 2-d plane

For each valid point, we use it's 'devectorized' form to plot the 90% region on the X-Y plane:



Drawbacks of FIM Approach

- A larger n (# samples used to calculate FIM) ($\gtrapprox 100$) gives a more 'stable' asymptotic covariance matrix and ellipses. We often do not have these many samples.
- With smaller n, data may not be representative of population, so we may get inaccurate results, such as above.
- For smaller n, plotting artifacts such as polygons may appear.

But...

Wishart Approach

The Wishart distribution is the distribution of sample covariance matrices for an IID sample drawn from a multivariate normal distribution. For the p-variate case with d ($\equiv n-1$) degrees of freedom, the probability density function (PDF) is:

$$f(\mathbf{X}) = \frac{1}{2^{\frac{dp}{2}} \det(\hat{\Sigma})^{\frac{d}{2}} \Gamma_p(\frac{d}{2})} \det(\mathbf{X})^{\frac{d-p-1}{2}} \cdot \exp\left(-\frac{1}{2}\mathrm{tr}(\hat{\Sigma}^{-1}\mathbf{X})\right),$$

where Γ_p is the p-variate gamma function defined as:

$$\Gamma_p\left(\frac{d}{2}\right) = \pi^{\frac{p(p-1)}{4}} \prod_{j=1}^{p} \Gamma\left(\frac{d}{2} - \frac{j-1}{2}\right).$$

Goal: Find the 90% Highest Density Region (HDR) for the Wishart PDF, then plot samples from the boundary of this region.

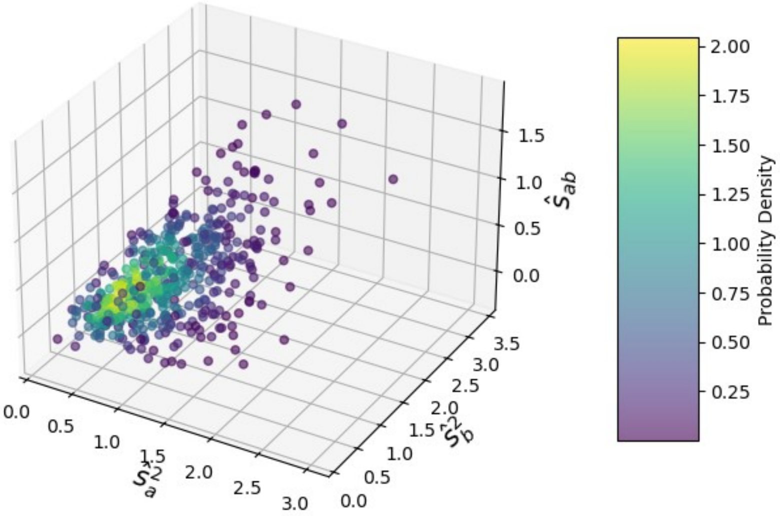
Plotting Wishart Samples

The samples drawn from the Wishart distribution are 2×2 covariance matrices. We vectorize these samples:

$$\begin{bmatrix} x & z \\ z & y \end{bmatrix} \mapsto \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3,$$

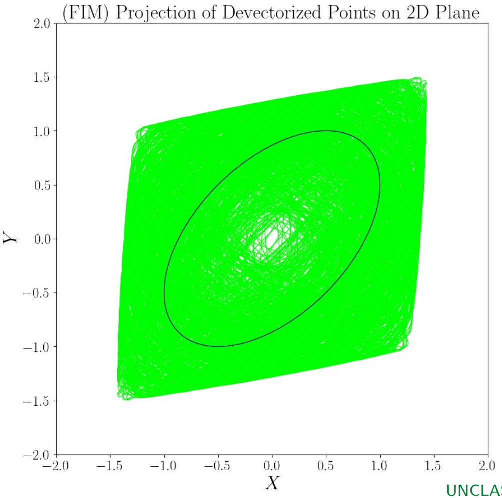
and plot them in 3-d. Lastly, we use a colorbar to indicate their PDF values.

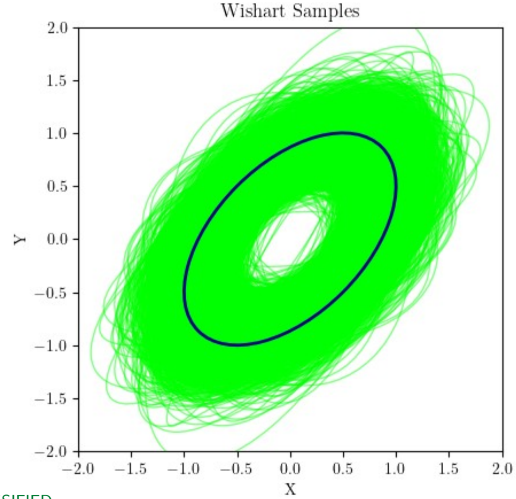
Plot of Wishart Samples



Comparing FIM and Wishart Projections

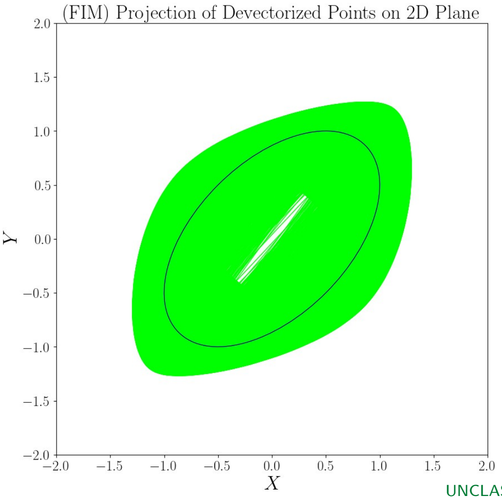
For $n = 10$:

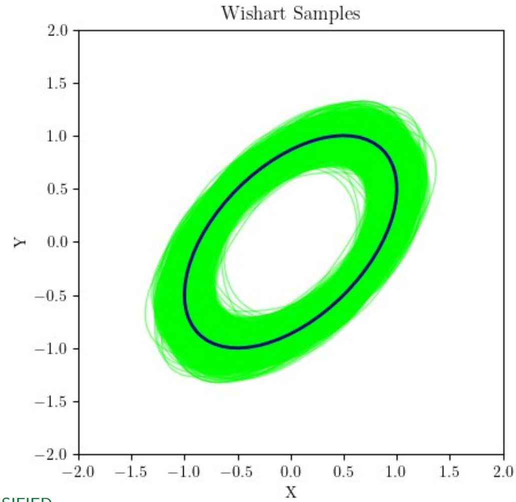




Comparing FIM and Wishart Projections

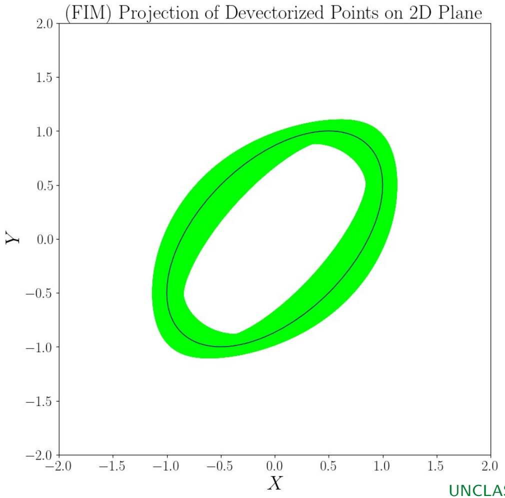
For $n = 50$:

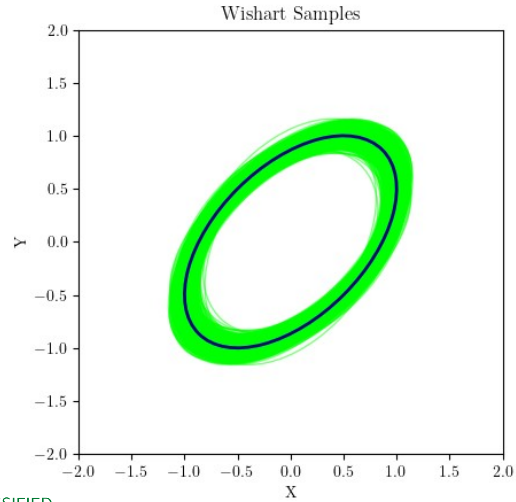




Comparing FIM and Wishart Projections

For $n = 200$:





Conclusion & Next Steps

Results:

- The FIM approach, which we currently use, may not be appropriate for small sample sizes.
- Using just the Wishart samples provides a tighter confidence region for the 1-σ ellipse, we should be able to tighten it further by taking the 90% HPD (and thus finding a true 90% confidence region).

Next steps: finding the boundary of the 90% HPD and plotting these points.

- Identify integration technique and solve for bounds.
- Sample points from boundary.
- Plot these points.

Thank you!

Acknowledgements

- Steven Reyes - Mentor
- Allan McQuarrie and Maximillian Chen - Valuable feedback on the project and presentation
- Daniel Lugar and Rachel Keil - Facilitating and helping with project #2
- Opel Jones and Christopher MacGahan - Guidance and help towards FPS intern challenge
- Matthew Villemarette - Helping with MLE concepts and resources
- Ash Riad - ATLAS mentor
- Fellow interns - in particular, our team for the FPS intern challenge: Robin Paranilam, Declan Willems, Savannah Burke, Dustin Sims

References

Thomas C. Henderson (2020)
How to Draw a Covariance Error Ellipse
https://users.cs.utah.edu/ tch/CS6640F2020/resources/How%20to%20draw...

Stanley H. Chan (2015)
ECE 645: Estimation Theory; Lecture 8: Properties of Maximum Likelihood Estimation (MLE)
https://engineering.purdue.edu/ChanGroup/ECE645Notes/StudentLecture0...

(FIM Approach) Choosing the scaling factor

On slide 13, we applied the transformation:

$$\mathbf{P} \mapsto \sqrt{\chi^2_{0.9,3}}\mathbf{LP}.$$

Why $\sqrt{\chi^2_{0.9,3}}$? The squared **Mahalanobis distance** $D^2 := (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)$ is equal to the squared norm of the standardized random vector $\mathbf{z} = \Sigma^{-\frac{1}{2}} (\mathbf{x} - \mu)$.

Given that $\mathbf{z}$ follows a standard normal distribution (i.e., $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$), the squared norm $||\mathbf{z}||^2 (= D^2)$ follows a chi-squared distribution with 3 degrees of freedom.

To find the 90% confidence region, we want to find c such that $\Pr(||\mathbf{z}||^2 \leq c) = 0.9$, which by definition is $\chi^2_{0.9,3}$.