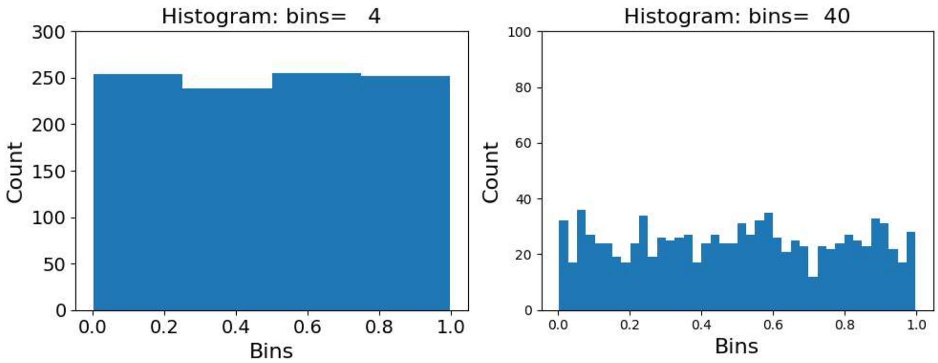
CO544 Machine Learning and Data Mining
E/19/432 Wickramaarachchi U.I.

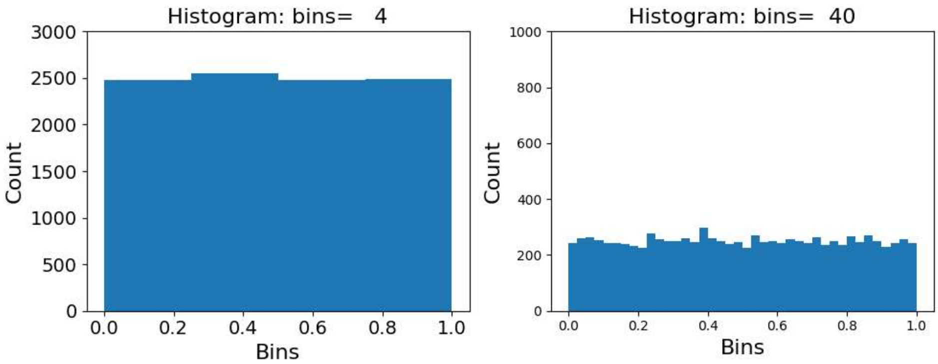
1. Random Numbers and Uni-variate Densities

Generate 1000 uniform random numbers and plot a histogram.

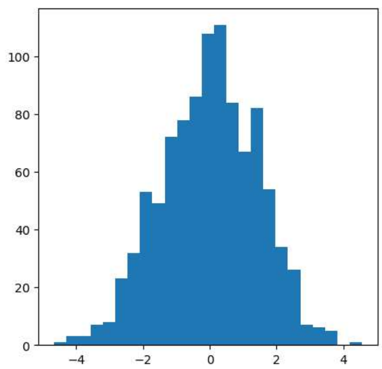


- Though the data is from a uniform distribution, the histogram does not appear flat. Why?
  The generate data is not uniformly ditributed in each bin.
- Every time you run it, the histogram looks slightly different? Why?
  The data is randomly generated. So, this does not assure that we get the same set of data everytime.
- How do the above observations change (if so how) if you had started with more data?
  The histrogram with less number of bins appear more uniformly ditributed.

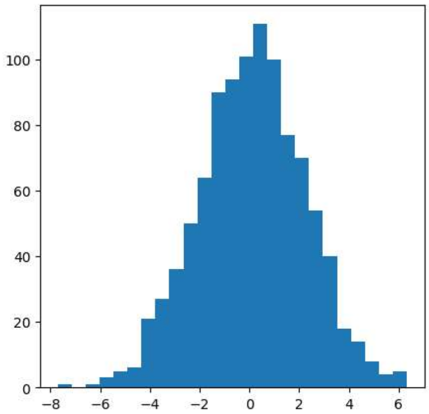
With 10,000 uniform random numbers



Add and subtract some uniform random numbers
What do you observe?
With N = 1000, 12 Random Numbers



How does the resulting histogram change when you change the number of uniform random numbers you add and subtract?
With N = 1000, 30 Random Numbers

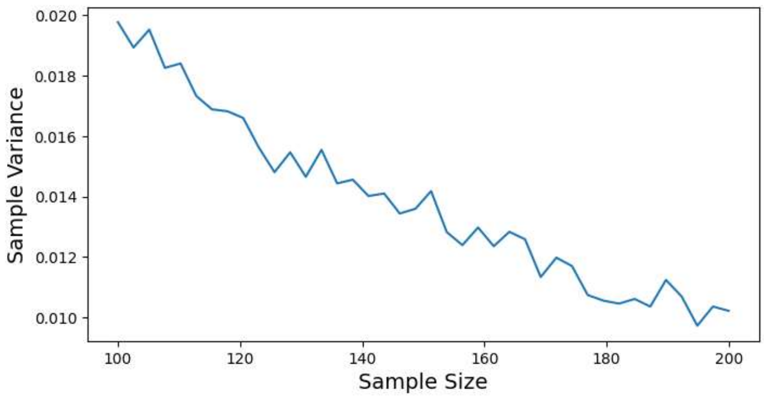


The distribution becomes closer to a gaussian distribuition

Is there a theory that explains your observation?
Central Limit Theorem

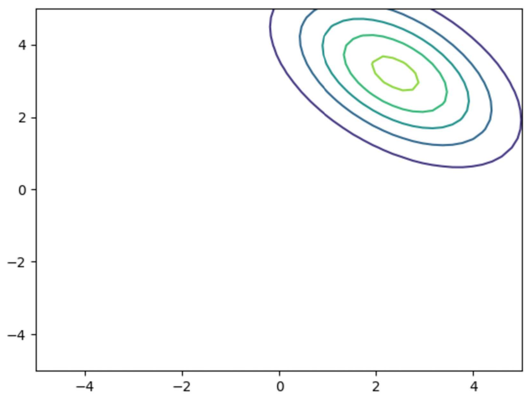
2. Uncertainty in Estimation

The following graph shows the sample variance with sample size. As we can see higher the sample size, lower the variance becomes.

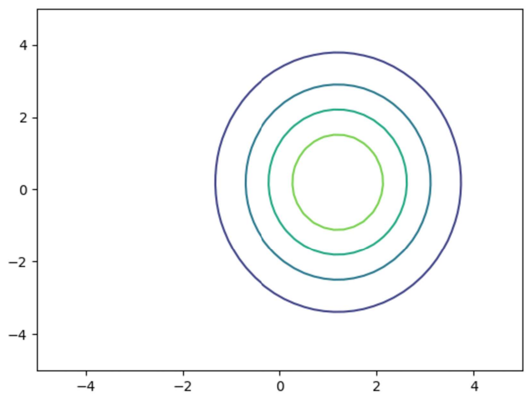


3. Bi-variate Gaussian Distribution

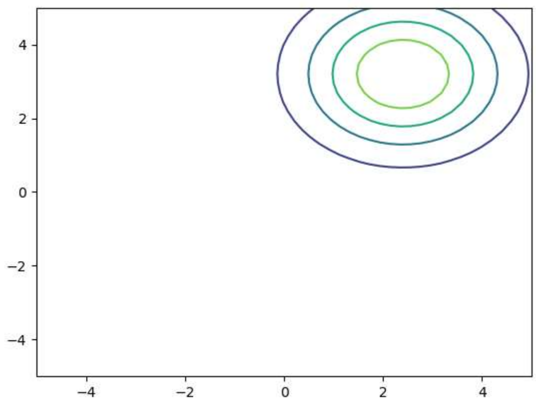
$\mathcal{N}\left(\begin{bmatrix} 2.4 \\ 3.2 \end{bmatrix}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\right)$



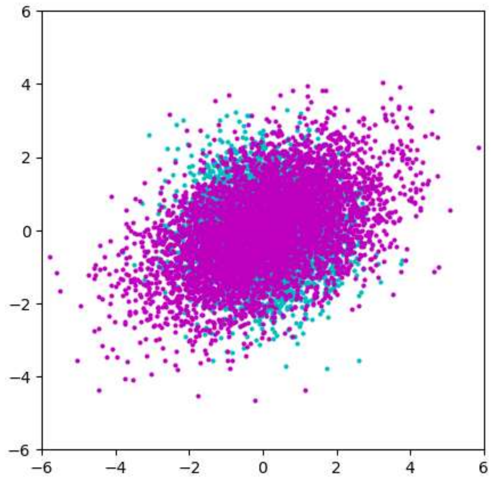
$\mathcal{N}\left(\begin{bmatrix} 1.2 \\ 0.2 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}\right)$



$$\mathcal{N}\left(\begin{bmatrix} 2.4 \\ 3.2 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\right)$$

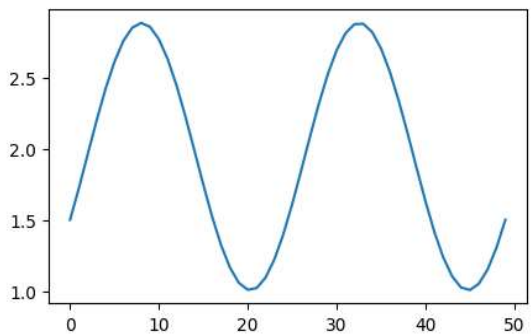


4. Sampling from a multi-variate Gaussian



5. Distribution of Projections

- The vector: [0.8660254037844386, 0.5000000000000001]
- Magnitude : 1.0
- Angle : 59.99999999999999
- Projected variance: 2.879677989361414



- What are the maxima and minima of the resulting plot?

  Minimum value: 1.01278486
  Maximum value: 2.88299609

- Compute the eigenvalues an eigenvectors of the covariance matrix C

  covariance_matrix = np.cov(Y.T)
  covariance_matrix= [[2.39051158 0.82371663] [0.82371663 1.50526937]]

- Eigenvalues of the covariance matrix:
  [2.88299609 1.01278486]

- Eigenvectors of the covariance matrix:
  [[0.85829426 -0.51315783]
  [0.51315783 0.85829426]]

- Can you see a relationship between the eignevalues and eigenvectors and the maxima and minima of the way the projected variance changes?
  The eigenvalues are similar to the maxima and minima values.

- The shape of the graph might have looked sinusoidal for this two dimensional problem. Can you analytically confirm if this might be true?
  Since the projected variance depends on the cosine of theta, and cos(theta) varies sinusoidally with theta, the plot of projected variance should exhibit a sinusoidal shape.