E9 205 Machine Learning for Signal Processing

ML, MAP, MMSE and Gaussian Modeling

04-09-2017

Recap …

- Decision Theory
    - Inference problem
        - Finding the joint density $p(\mathbf{x}, \mathbf{t})$
    - Decision problem
        - Using the inference to make the classification or regression decision

Decision Problem - Classification

- Minimizing the mis-classification error
- Decision based on maximum posteriors

$$argmax_j \; p(C_j|\mathbf{x})$$

- Loss matrix
  - Minimizing the expected loss

$$argmax_j \; \sum_k L_{k,j} p(C_k|\mathbf{x})$$

Approaches for Inference and Decision

I. Finding the joint density from the data.

$$p(C_k|\mathbf{x}) \; \alpha \; p(\mathbf{x}|C_k)p(C_k)$$

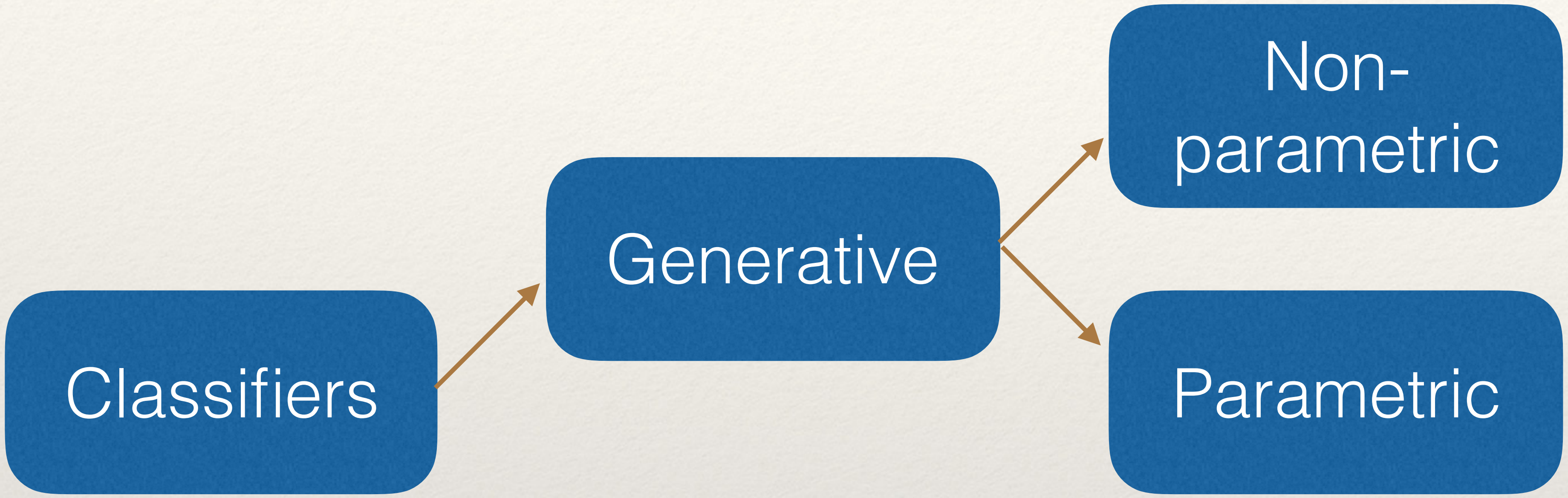
II. Finding the posteriors directly.

III. Using discriminant functions for classification.

Decision Rule for Regression

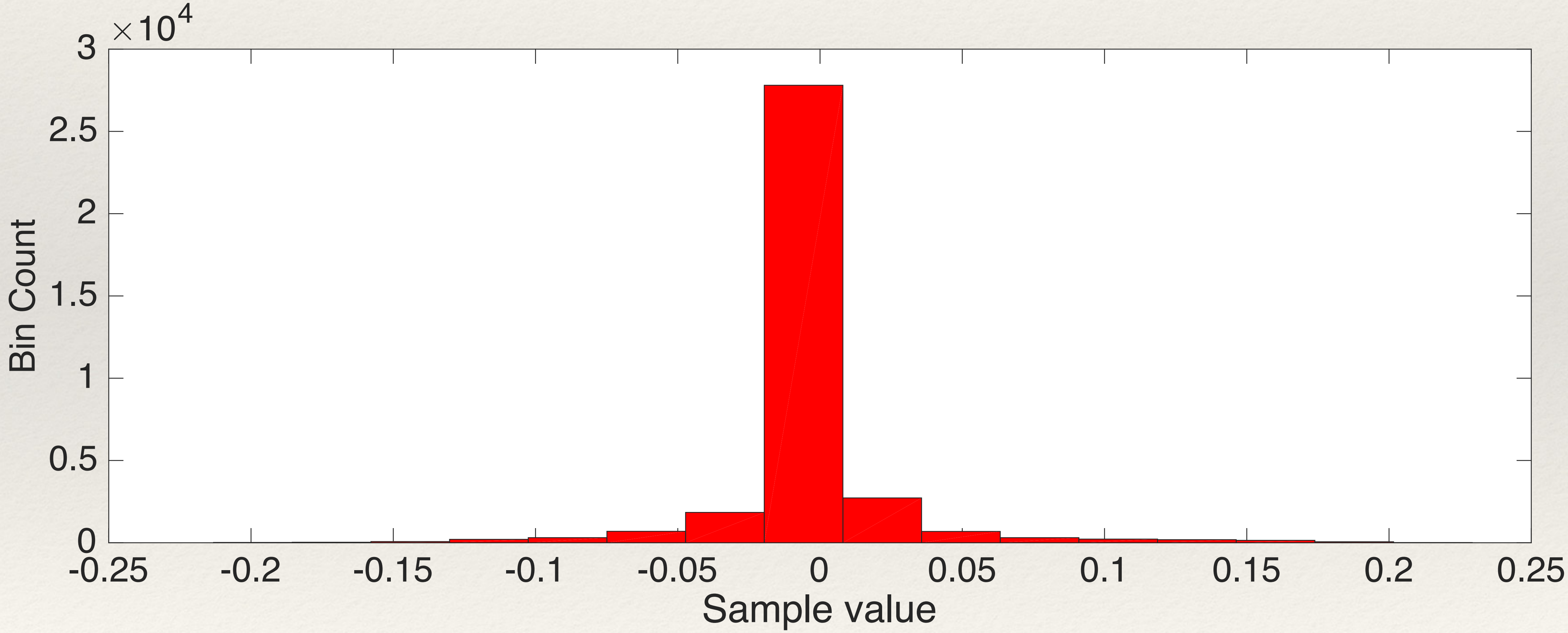
- Minimum mean square error loss
- Solution is conditional expectation.

Generative Modeling



Non-parametric Modeling

- Non-parametric models do not specify an apriori set of parameters to model the distribution. Example - Histogram

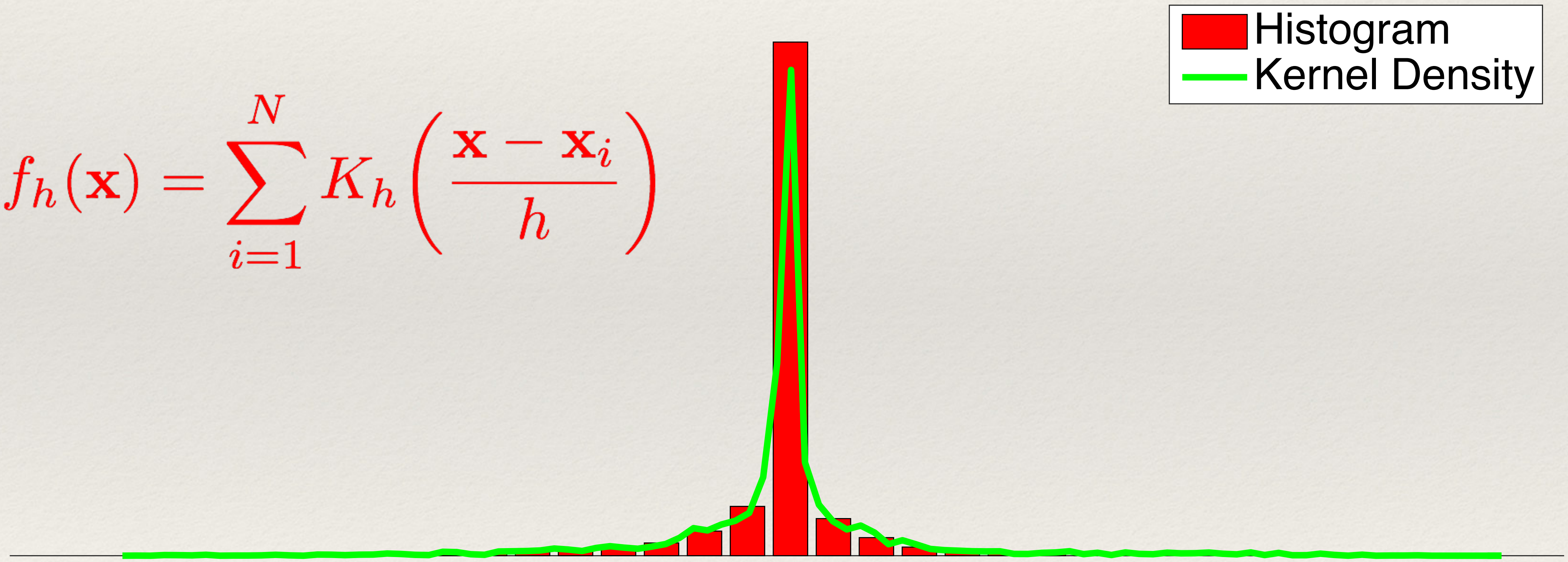


The density is not smooth and has block like shape.

Non-parametric Modeling

- Non-parametric models do not specify an apriori set of parameters to model the distribution.
  - Example - Kernel Density Estimators

$$f_h(\mathbf{x}) = \sum_{i=1}^{N} K_h\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right)$$



Kernel is a smooth function which obeys certain properties

Non-parametric Modeling

- Non-parametric methods are dependent on number of data points
  - Estimation is difficult for large datasets.
- Likelihood computation and model comparisons are hard.
- Limited use in classifiers

Parametric Models

- Collection of probability distributions which are described by a finite dimensional parameter set

$$\boldsymbol{\theta} = (\theta_1, \theta_2, ...\theta_K) \qquad P = \{P_{\boldsymbol{\theta}}\}$$

- Examples -
  - Poisson Distribution

$$p_\lambda(j) = \frac{\lambda^j}{j!}e^{-\lambda}$$

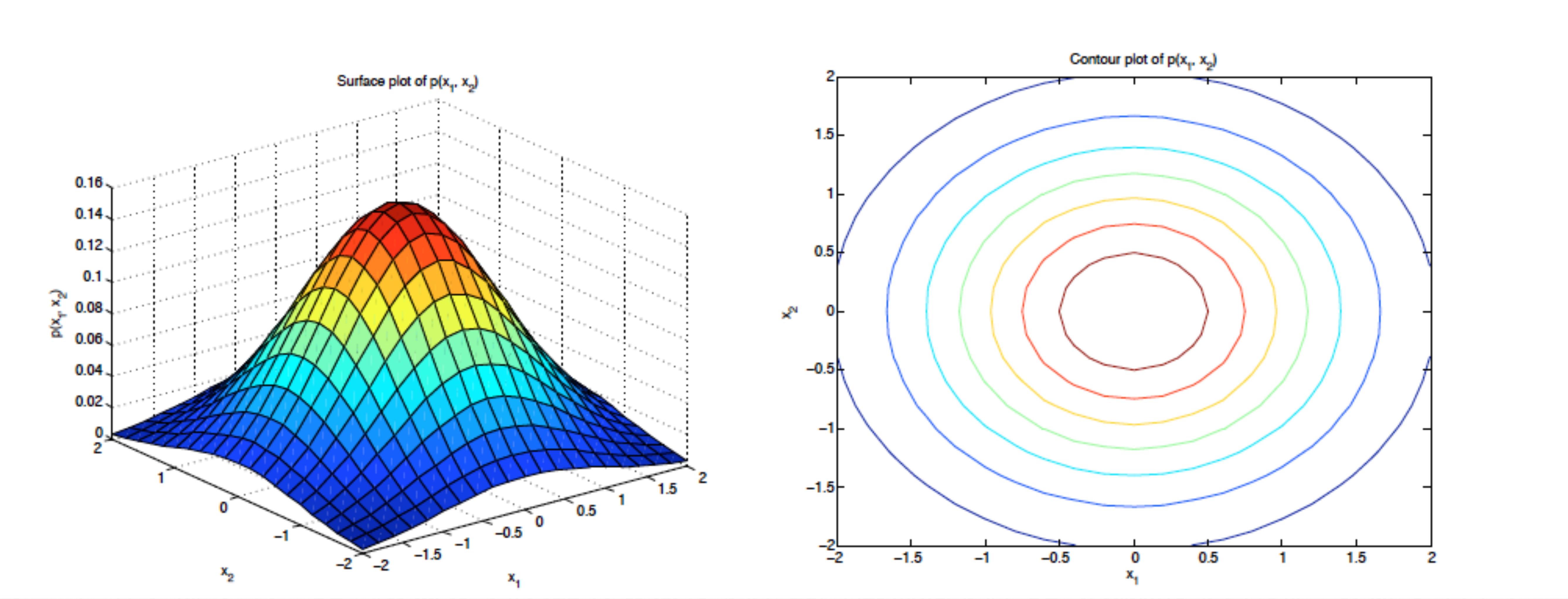
  - Bernoulli Distribution

$$p(\mathbf{x}|\boldsymbol{\mu}) = \prod_{i=1}^{D} \mu_i^{x_i}(1-\mu_i)^{x_i}$$

  - Gaussian Distribution

$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^D|\boldsymbol{\Sigma}|}} exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^*\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

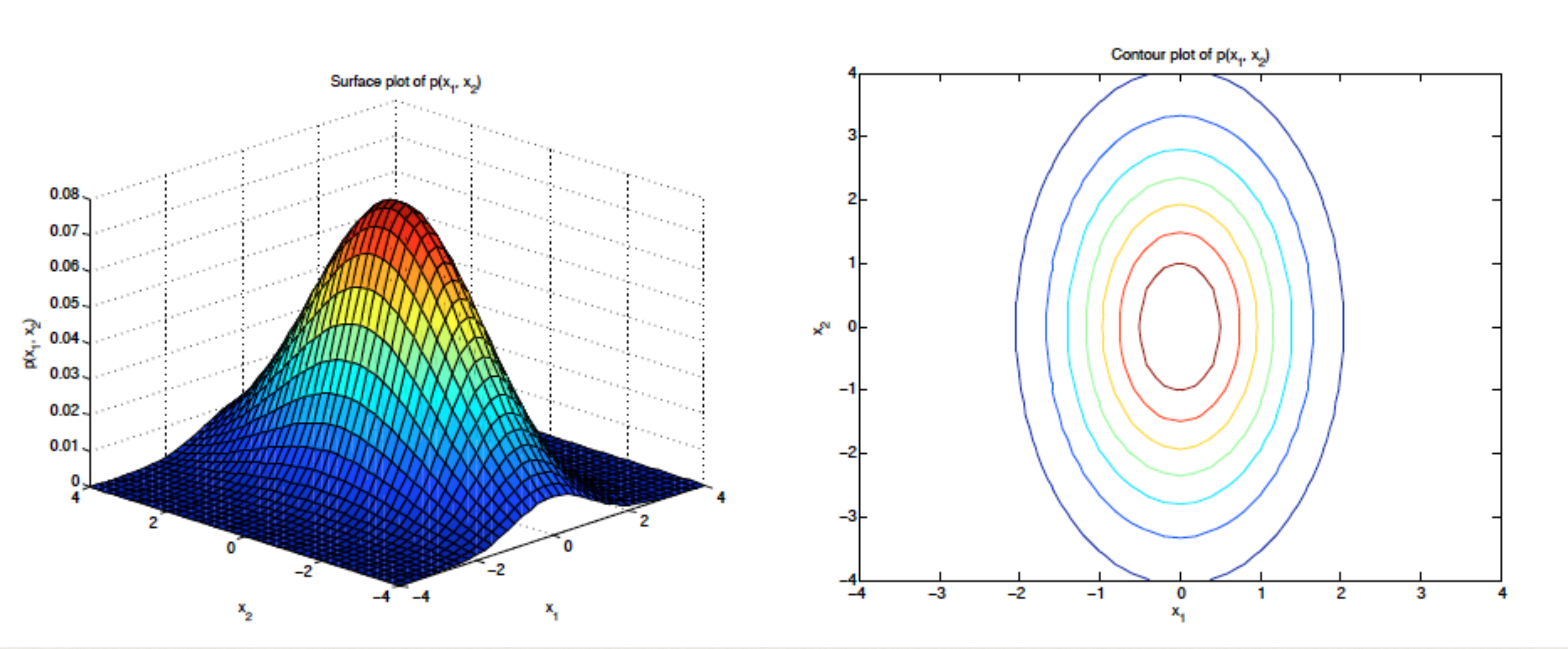
Gaussian Distribution



Points of equal probability lie on on contour

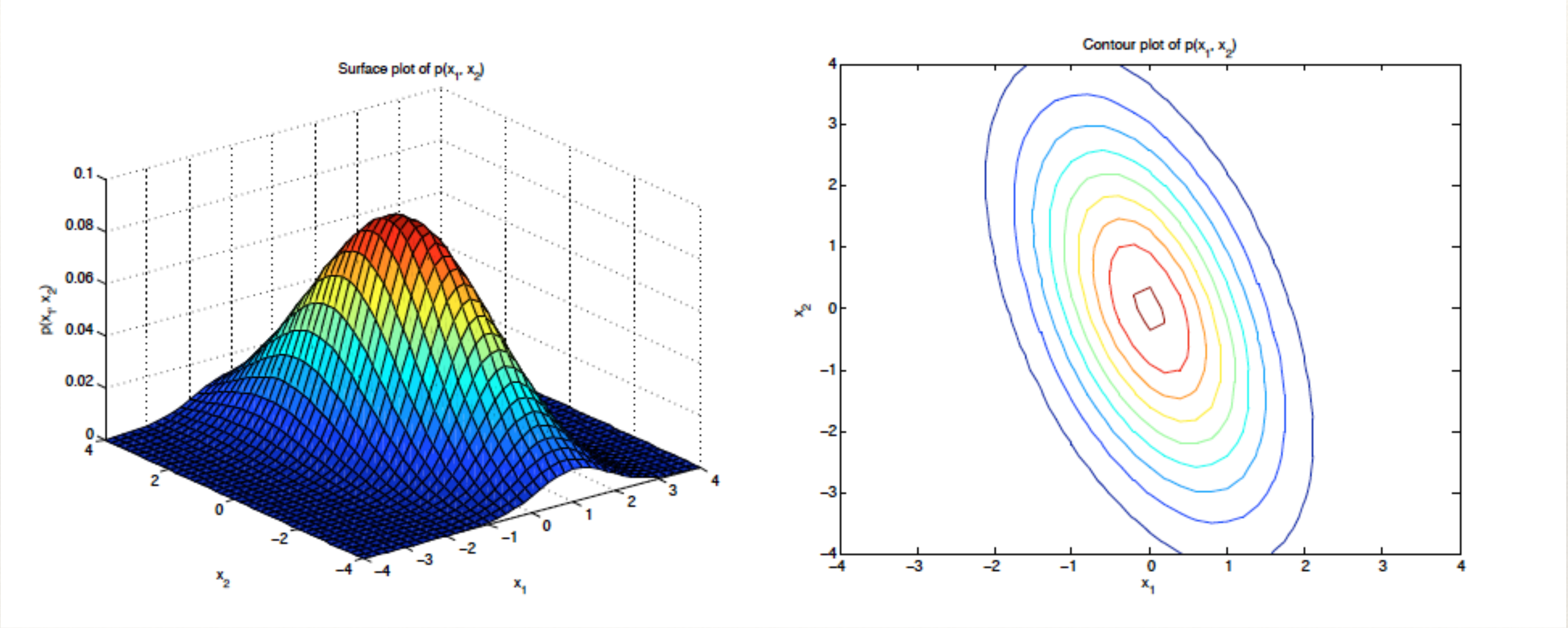
Diagonal Gaussian with Identical Variance

Gaussian Distribution



Diagonal Gaussian with different variance

Gaussian Distribution



Full covariance Gaussian distribution