The Multivariate Gaussian Distribution

Analysis of Background Magnetoencephalogram Noise

Courtesy of Simona Temereanca MGH Martinos Center for Biomedical Imaging
Magnetoencephalogram Background Noise

Boxplot  Q-Q Plot

Why are these data Gaussian? Answer: Central Limit Theorem

1.6 Seconds of Two Time Series of MEG Recordings

Front Sensor

Back Sensor

Figure 4G
Figure 4H

Scatterplot of MEG Recordings

Figure 4I

Histograms and Q-Q Plots of MEG Recordings

Front Sensor

Back Sensor
Case 2: Probability Model for Spike Sorting

The data are tetrode recordings (four electrodes) of the peak voltages (mV) corresponding to putative spike events from a rat hippocampal neuron recorded during a texture-sensitivity behavioral task.

Each of the 15,600 spike events recorded during the 50 minutes is a four vector.

The objective is to develop a probability model to describe the cluster of spikes events coming from a single neuron.

Such a model provides the basis for a spike sorting algorithm.

Acknowledgments: Data provided by Sujith Vijayan and Matt Wilson

Technical Assistance Julie Scott
Time-Series Plot of Tetrode Recordings

Channel Number

Six Bivariate Plots of Tetrode Channel Recordings
Histograms of Spike Events By Channel

Box Plots of Spike Events By Channel
DATA: The Tetrode Recordings

\[ x_k = \begin{bmatrix} x_{k,1} \\ x_{k,2} \\ x_{k,3} \\ x_{k,4} \end{bmatrix} \]

Four peak voltages recorded on the k-th spike event for \( k = 1, \ldots, K \), where \( K \) is the total number of spike events.

GAUSSIAN PROBABILITY MODEL

Four-Variate Gaussian Model

\[
f(x_k | \mu, W) = \frac{1}{(2\pi)^2 |W|^\frac{1}{2}} \exp \left\{ -\frac{1}{2} (x_k - \mu)^\top W^{-1} (x_k - \mu) \right\}
\]

Mean

\[
\mu = (\mu_1, \mu_2, \mu_3, \mu_4)
\]

Covariance Matrix (symmetric)

\[
W = \begin{bmatrix}
\sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\
\sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\
\sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} \\
\sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2
\end{bmatrix}
\]

\( k = 1, \ldots, K \)
N-Multivariate Gaussian Model Factoids

1. A Gaussian probability density is completely defined by its mean vector and covariance matrix.

2. All marginal probability densities are univariate Gaussian.

3. Frequently used because it is
   i) analytically and computationally tractable
   ii) suggested by the Central Limit Theorem

4. Any linear of the components is Gaussian (a characterization).

Central Limit Theorem

The distribution of the sum of random quantities such that the contribution of any individual quantity goes to zero as the number of quantities being summed becomes large (goes to infinity) will be Gaussian.
**N-Multivariate Gaussian Model Factoids**

Bivariate Gaussian Distribution

Cross-section is an ellipse

Marginal distribution is univariate Gaussian

Cumulative Distribution Function

**Univariate Gaussian Model Factoids**

Gaussian Probability Density Function

\[ f(x) = (2\pi \sigma^2)^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2} \right\}. \]

Standard Gaussian Probability Density Function

\[ f(x) = (2\pi)^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} x^2 \right\}. \]

\[ \mu = 0 \quad \sigma^2 = 1 \]

Standard Cumulative Gaussian Distribution Function

\[ \Phi(x) = \int_{-\infty}^{x} (2\pi)^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} u^2 \right\} du. \]
Univariate Gaussian Model Factoids

Mu is the mean (location)

Standard deviation (scale)

Any Gaussian distribution can be converted into a standard Gaussian distribution (mu = 0, sd =1)

68% of the area within ~ 1 sd of mean

95% of the area within ~ 2 sd of mean

99% of the area within ~ 2.58 sd of mean

ESTIMATION

Joint Distribution of the Four-Variate Gaussian Model

\[
f(x | \mu, W) = \prod_{k=1}^{K} f(x_k | \mu, W) = \left( \frac{1}{(2\pi)^2 |W|^{\frac{1}{2}}} \right)^K \exp \left\{ -\frac{1}{2} \sum_{k=1}^{K} (x_k - \mu)W^{-1}(x_k - \mu) \right\}
\]

where \( x = (x_1, ..., x_K) \)

Log Likelihood

\[
\log f(x | \mu, W) = -K\log(2\pi) - \frac{K}{2} \log |W| - \frac{1}{2} \sum_{k=1}^{K} (x_k - \mu)^TW(x_k - \mu)
\]

where K is the number of spike events in the data set.
ESTIMATION

For Gaussian observations the maximum likelihood and method-of-moments estimates are the same.

\[ \hat{\mu}_i = K^{-1} \sum_{k=1}^{K} x_{k,i} \]
\[ \hat{\sigma}_i^2 = K^{-1} \sum_{k=1}^{K} (x_{k,i} - \hat{\mu}_i)^2 \]

Sample Covariance
\[ \hat{\sigma}_{i,j} = K^{-1} \sum_{k=1}^{K} (x_{k,i} - \hat{\mu}_i)(x_{k,j} - \hat{\mu}_j) \]
Sample Correlation
\[ \hat{\rho}_{i,j} = \frac{\hat{\sigma}_{i,j}}{\sqrt{\hat{\sigma}_{i,i} \hat{\sigma}_{j,j}}} \]
for i = 1, … , 4 and j = 1, … , 4.

CONFIDENCE INTERVALS FOR THE PARAMETER ESTIMATES OF THE MARGINAL GAUSSIAN DISTRIBUTIONS

The Fisher Information Matrix is
\[ I(\theta) = -E \left( \frac{\partial^2 L}{\partial \theta^2} \right) \]
\[ I(\theta) = \begin{bmatrix} K/\sigma^2 & 2K/\sigma^4 \\ 2K/\sigma^4 & 2K/\sigma^4 \end{bmatrix} \]
where \( \theta = (\mu_i, \sigma_i^2) \)

The confidence interval is
\[ \theta_{ii} \pm z_{\alpha/2} \sqrt{I(\theta)_{ii}^{-1}} \]
Four-Variate Gaussian Model Parameter Estimates

Sample Mean Vector

\[
\begin{bmatrix}
0.0015 \\
0.0019 \\
0.0011 \\
0.0009
\end{bmatrix}
\]

Sample Covariance Matrix

\[
\begin{bmatrix}
0.2322 & 0.1724 & 0.1503 & 0.1570 \\
0.1724 & 0.2304 & 0.1560 & 0.1387 \\
0.1503 & 0.1560 & 0.2126 & 0.1466 \\
0.1570 & 0.1387 & 0.1466 & 0.2130
\end{bmatrix}
\]

Sample Correlation Matrix

\[
\begin{bmatrix}
1.00 & 0.74 & 0.68 & 0.71 \\
0.74 & 1.00 & 0.70 & 0.63 \\
0.68 & 0.70 & 1.00 & 0.69 \\
0.71 & 0.63 & 0.69 & 1.00
\end{bmatrix}
\]

Marginal Gaussian Parameter Estimates and Confidence Intervals
(An Exercise: Compute the Confidence Intervals)

Sample Mean Vector

\[
\begin{bmatrix}
0.0015 \\
0.0019 \\
0.0011 \\
0.0009
\end{bmatrix}
\]

Sample Variances

\[
\begin{bmatrix}
0.2322 \\
0.2304 \\
1.0 \times 10^{-7} x \\
0.2126 \\
0.2130
\end{bmatrix}
\]
Histograms of Spike Events By Channel

Channel 1

Channel 2

Channel 3

Channel 4

Empirical and Model Estimates of Marginal Cumulative Probability Densities
GOODNESS-OF-FIT

• Q-Q Plots

• Kolmogorov-Smirnov Tests

• A Chi-Squared Test
  Separate the bivariate data into deciles and compute
  \[ \chi^2 \sim \sum_{d=1}^{10} \frac{(O_i - E_i)^2}{O_i} \]
  where \( O_i \) is the observed number of observation in decile \( i \) and \( E_i \) is expected number of observations in decile \( i \).
Q-Q Plots

Channel 1

Channel 2

Channel 3

Channel 4

K-S Plots

Channel 1

Channel 2

Channel 3

Channel 4
Six Bivariate Plots of Tetrode Channel Recordings

With 95% Probability Contour

Bivariate Plots of Tetrode Channel Recordings

Channel 4 vs 1  Channel 3 vs 2
Bivariate Plots of Tetrode Channel Recordings with Gaussian Equiprobability Contours

(An Exercise: Carry Out the Chi-Squared Test)

Channel 4 vs 1  Channel 3 vs 2

Linear Combinations of Gaussian Random Variables are Gaussian

If

$$X \sim N(\mu, W)$$

where

$$X = (x_1, x_2, x_3, x_4)$$

and

$$Y = \sum_{i=1}^{4} c_i x_i$$

where

$$c = (c_1, c_2, c_3, c_4)$$

then

$$Y \sim N(\sum_{i=1}^{4} c_i \mu_i, c^t W c)$$
CONCLUSION

• The data seem well approximated with a four-variate Gaussian model.

• The marginal probability density of Channel 4 is the best Gaussian fit.

The Central Limit Theorem most likely explains why the Gaussian model works here.
Epilogue

• Another real example of real Gaussian data in neuroscience data?