

# Combination of Multivariate Gaussian Distributions through Error Ellipses

Oktay Erten<sup>1</sup> and Clayton V. Deutsch<sup>2</sup>

<sup>1</sup>University of Alberta

<sup>2</sup>University of Alberta

## Learning Objectives

- Understand how to draw an error ellipse (or density contour) of a bivariate Gaussian distribution considering a given confidence level.
- Derive the equations for combining multivariate Gaussian distributions through error ellipses.
- Appreciate the error ellipses method using a simulated bivariate Gaussian dataset.

## 1 Introduction

The combination of probability distributions is a common challenge encountered in geology, engineering, and statistics. The distributions come from different data sources or sensors. The distributions could be univariate, bivariate or multivariate. The techniques to combine the distributions include (1) formulating and parameterizing a high dimensional distribution from which conditional distributions can be computed - called cokriging in geostatistics (C. V. Deutsch & Journel, 1998; Doyen, 1988) (2) considering a known prior distribution and making a neutral assumption about the redundancy in the information content - called collocated cokriging or Bayesian Updating in geostatistics (Doyen, Boer, & Pillet, 1996; Xu, Tran, Srivastava, & Journel, 1992), and (3) considering no prior distribution and conditional independence in the likelihood distributions - called Error Ellipses in the literature (Blachman, 1989). This Lesson aims to develop the latter technique.

There is a reasonably informed prior distribution in most spatial prediction problems; it comes from a careful choice of a stationary population and declustering of the available data. The error ellipses method would not be appropriate. There are situations, however, where there is no prior and the data sources are very different; the error ellipses technique is useful when the convex estimation of the probability distribution is required. For example, the horizontal variogram calculated from a few wells are highly uncertain, whereas the horizontal variogram calculated from dense seismic information would be more structured but noisier. The estimation of the horizontal variogram from the wells can be improved by combining the distributions of the squared differences from wells and seismic information for each lag (Rezvandehey, 2016). In this Lesson, the construction of the confidence ellipse of a bivariate Gaussian distribution is first reviewed. The error ellipse method and derivation of its equations are explained, and the method is demonstrated using a simulated bivariate Gaussian dataset.

## 2 Multivariate Gaussian Distribution

### Probability Density Function

Consider a  $p \times 1$  random vector  $\mathbf{X} = [X_1, \dots, X_p]^T$  that has a multivariate Gaussian distribution  $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ . Its probability density function is given as follows:

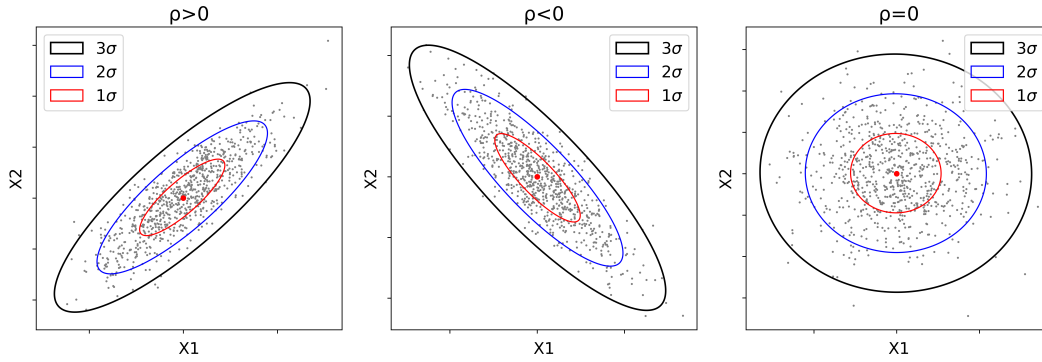


Figure 1: Confidence ellipses representing different confidence levels

$$f(\mathbf{X}|\boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu}) \right\}$$

where  $p$  is the dimension of the vector;  $\boldsymbol{\mu}$  is the  $p \times 1$  vector of means  $\boldsymbol{\mu} = [\mu_1, \dots, \mu_p]^T$ , and  $\Sigma$  is the  $p \times p$  positive definite covariance matrix  $\Sigma$  of  $\mathbf{X}$ :

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

If a random vector  $\mathbf{X}$  is multivariate Gaussian, each random variable of  $\mathbf{X}$  is also Gaussian. If  $\rho = 0$ , that is,  $\Sigma$  is of a diagonal matrix (all off-diagonal elements are zero), the multivariate Gaussian distribution is equal to  $p$  number of univariate Gaussian distributions (Johnson & Wichern, 1988).

### 3 Density Contours of a Bivariate Gaussian Distribution

The plot of the density (or probability) contours of a bivariate Gaussian distribution represents a three-dimensional surface. The constant probability contours, however, can be plotted on a two-dimensional format by considering the same height on the z-axis (or the constant height of the surface). Consider that a random vector  $\mathbf{X} = [X_1, X_2]$  has a bivariate ( $p = 2$ ) Gaussian distribution. The scalar quantity of the square of the distance  $(\mathbf{X} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu}) = K$  is referred to as the Mahalanobis distance of the vector  $\mathbf{X}$  to the mean vector  $\boldsymbol{\mu}$ . The surface on which the random variable  $K$  is constant is an ellipse (or a  $p$ -dimensional ellipsoid in the multivariate case) centered at  $\boldsymbol{\mu} = [\mu_1, \mu_2]$ . This ellipse (or a probability contour) defines the region of a minimum area (or volume in multivariate case) containing a given probability under the Gaussian assumption. Figure 1 illustrates confidence (error) ellipses with different confidence levels (i.e. 68%, 90%, 95%), considering the cases where the random variables are (1) positively correlated  $\rho > 0$ , (2) negatively correlated  $\rho < 0$ , and (3) independent  $\rho = 0$ .

This figure shows how the ranges (or lengths) and directions of the axes of the ellipses change depending on the selected confidence level and the covariance matrix of the random vector  $\mathbf{X}$ . The confidence ellipses constructed based on the given confidence levels can be used to check the bivariate Gaussianity of a given distribution (J. L. Deutsch & Deutsch, 2011).

### Step 1: Selecting a probability level for the error ellipse

The quadratic form defining the scalar random variable  $K$  has a chi-square distribution  $\chi_2^2(\alpha)$  with two degrees of freedom:

$$P\{K \leq \chi_2^2(\alpha)\} = P\{(\mathbf{X} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu}) \leq \chi_2^2(\alpha)\} = 1 - \alpha$$

where  $\chi_2^2(\alpha)$  is the upper  $(100\alpha)\text{th}$  percentile of the  $\chi^2$ -distribution with two degrees of freedom.  $1 - \alpha$  is the probability (or a confidence level) that the value of the random vector  $\mathbf{X}$  will be inside the ellipse. For example, consider that the given probability is  $\alpha = 0.05$ , that is,  $\chi_2^2(0.05) = 5.99$ , then the 95% confidence ellipse is defined by:

$$P\{K \leq \chi_2^2(0.05)\} = P\{(\mathbf{X} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu}) \leq 5.99\} = 0.95$$

### Step 2: Calculating the eigenvalues of the covariance matrix

The positive definite ( $|\Sigma| \neq 0$  and  $\Sigma^{-1}$  exists) covariance matrix  $\Sigma$  of the random vector  $\mathbf{X} = [X_1, X_2]$  is as follows:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

The eigenvalues  $\boldsymbol{\lambda} = [\lambda_1, \lambda_2]$  of  $\Sigma$  are calculated by:

$$\lambda_1 = \frac{1}{2} \left[ \sigma_{11} + \sigma_{22} + \sqrt{(\sigma_{11} - \sigma_{22})^2 + 4\sigma_{11}\sigma_{22}\rho^2} \right] \quad \lambda_2 = \frac{1}{2} \left[ \sigma_{11} + \sigma_{22} - \sqrt{(\sigma_{11} - \sigma_{22})^2 + 4\sigma_{11}\sigma_{22}\rho^2} \right]$$

where  $\sigma_{11}$  and  $\sigma_{22}$  are the variances of the random variables  $X_1$  and  $X_2$ , and  $\rho$  is the linear correlation coefficient.

### Step 3: Calculating the lengths of the ellipse axes

The lengths of the ellipse axes are a function of the given probability  $\chi_2^2(\alpha)$ , the eigenvalues  $\boldsymbol{\lambda} = [\lambda_1, \lambda_2]$  and the linear correlation coefficient  $\rho$ . For example, 95% confidence ellipse is defined by:

$$[X_1 - \mu_1 \quad X_2 - \mu_2] \Sigma^{-1} \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{bmatrix} \leq \chi_2^2(0.05)$$

As  $\Sigma$  denotes a symmetric matrix and  $\sigma_{11} \neq \sigma_{22}$ , the eigenvectors of  $\Sigma$  is linearly independent (or orthogonal). Therefore,  $\Sigma$  can be written as follows:

$$\Sigma = T D T^{-1}$$

where  $T = [v_1 | v_2]$  are the eigenvectors of  $\Sigma$  and  $D$  is the diagonal matrix of the eigenvalues  $\boldsymbol{\lambda} = [\lambda_1, \lambda_2]$ :

$$D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

replacing  $\Sigma^{-1}$  by  $T D^{-1} T^{-1}$ , the square of the difference can be written as:

$$[X_1 - \mu_1 \quad X_2 - \mu_2] T D^{-1} T^{-1} \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{bmatrix} \leq \chi_2^2(0.05)$$

Denoting

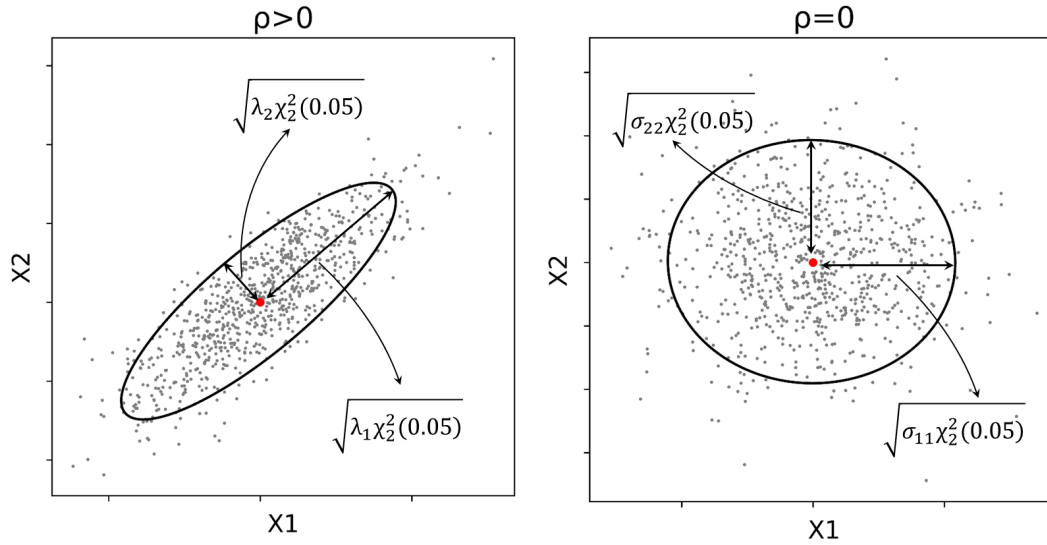


Figure 2: 95% confidence ellipses and their semi-diameters when  $\rho > 0$  and  $\rho = 0$

$$\begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} = T^{-1} \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{bmatrix}$$

As  $T^T = T^{-1}$ , the square of difference can be expressed as:

$$[\omega_1 \ \omega_2] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}^{-1} \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} \leq \chi^2_2(0.05)$$

If the above equation is further evaluated, The resulting equation is the equation of an ellipse aligned with the axis  $\omega_1$  and  $\omega_2$  in the new coordinate system.

$$\frac{\omega_1^2}{\chi^2_2(0.05)\lambda_1} + \frac{\omega_2^2}{\chi^2_2(0.05)\lambda_2} \leq 1$$

The axes of the ellipse are defined by  $\omega_1$  axis with a length  $2\sqrt{\chi^2_2(0.05)\lambda_1}$  and  $\omega_2$  axis with a length  $2\sqrt{\chi^2_2(0.05)\lambda_2}$ .

When  $\rho = 0$ , the eigenvalues are equal to  $\lambda_1 = \sigma_{11}$  and  $\lambda_2 = \sigma_{22}$ . Also,  $T$  matrix whose elements are the eigenvectors of  $\Sigma$  becomes an identity matrix. The final equation of an ellipse is then defined by:

$$\frac{(X_1 - \mu_1)^2}{\chi^2_2(0.05)\lambda_1} + \frac{(X_2 - \mu_2)^2}{\chi^2_2(0.05)\lambda_2} \leq 1$$

It is clear from the equation given above that the axes of the ellipse are parallel to the coordinate axes. The lengths of the axes of the ellipse are then defined as  $2\sqrt{\sigma_{11}\chi^2_2(0.05)}$  and  $2\sqrt{\sigma_{22}\chi^2_2(0.05)}$ . Figure 2 displays 95% confidence ellipses drawn when the variables are positively correlated and when they are independent.

It is clearly seen from Figure 2 that when  $\rho > 0$ , the axes of the ellipse are aligned with the rotated axes in the transformed coordinate system, and when  $\rho = 0$ , axes of the ellipse are parallel to the original coordinate system.

## 4 Error Ellipses Method

The error ellipses method, first introduced by (Blachman, 1989), aims to combine individual probability distributions by the linear weighted averaging of the means. The optimal weighting minimizes the resulting variance of the combined distribution:

$$C_{\bar{x}} = \left[ \sum_{i=1}^n C_i^{-1} \right]^{-1}$$

where  $C_{\bar{x}}$  is the minimized covariance matrix of the combined probability distribution, and  $C_i^{-1}$  is the covariance matrix of the individual probability distributions. The estimated mean vector of the combined distribution is then defined by:

$$\bar{x} = C_{\bar{x}} \left[ \sum_{i=1}^n C_i^{-1} x_i \right]$$

where  $\bar{x}$  is the mean of the combined probability distribution.

### Derivation of the Error Ellipse Equations-1D Case

Consider that  $\{x_i; i = 1, \dots, n\}$  are the means of  $n$  number of probability distributions, and  $\{\sigma_i^2; i = 1, \dots, n\}$  are the corresponding variances of these probability distributions. The unbiased estimate of the mean of the combined probability distribution is defined by:

$$\bar{x} = \frac{\sum_{i=1}^n \lambda_i x_i}{\sum_{i=1}^n \lambda_i}$$

The variance of the estimate  $\bar{x}$  can be written as:

$$Var[\bar{x}] = \sum_{i=1}^n \left[ \lambda_i / \sum_{i=1}^n \lambda_i \right]^2 Var[x_i]$$

Recall that  $Var[\bar{x}] = E[\bar{x}^2] - \{E[\bar{x}]\}^2$ . Differentiating the equation given above with respect to  $\lambda_j$  guarantees the condition for the variance to be minimum:

$$\lambda_j Var[\bar{x}] = \frac{\sum_{i=1}^n \lambda_i^2 Var[x_i]}{\sum_{i=1}^n \lambda_i}$$

The right hand side of the equation does not depend on  $j$  and therefore,  $\lambda_j$  is proportional to  $1/Var[\bar{x}]$ . Substituting these weights into the first and second equations given above yields:

$$\bar{x} = \left[ \sum_{i=1}^n Var[\bar{x}]^{-1} \right]^{-1} \sum_{i=1}^n Var[\bar{x}]^{-1} x_i$$

$$Var[\bar{x}] = \left[ \sum_{i=1}^n Var[\bar{x}]^{-1} \right]^{-1}$$

Finally, the mean of the combined distribution with the minimized variance can then be defined by:

$$\bar{x} = Var[\bar{x}] \sum_{i=1}^n Var[\bar{x}_i]^{-1} \bar{x}_i$$

### Derivation of the Error Ellipse Equations in a Multivariate Case

Consider  $m$  independent sets of observations,  $\mathbf{O} = [O_1, \dots, O_m]$  of the same target. Each observation  $O_i, i = 1, \dots, m$  is an individual estimate of the target. The error ellipse of each observation is a contour of the Gaussian conditional probability density function  $P(x, y|O_i)$  of the target (Blachman, 1989). Considering the Bayes' theorem, for which the readers could refer to the lesson by (C. V. Deutsch & Deutsch, 2018), the aforementioned posterior distribution based on one set of observation can be defined by:

$$P(x, y|O_i) = \frac{P_i(x, y) * P(O_i|x, y)}{P(O_i)}$$

Provided that  $P_i(x, y)$  is constant over the area of interest, the posterior distribution can be defined by:

$$P(O_i|x, y) = \frac{P(O_i) * P(x, y|O_i)}{P_i(x, y)}$$

Assuming that each posterior distribution  $P(O_i|x, y)$  is independent, the total conditional probability over all observations,  $P(O_1, \dots, O_m|x, y)$  can then be defined as the product of individual posterior distributions  $P(O_i|x, y)$ , that is,

$$P(O_1, \dots, O_m|x, y) = \prod_{i=1}^m \frac{P(O_i) * P(x, y|O_i)}{P_i(x, y)}$$

The conditional probability of the target given all observations is defined by the following function:

$$P(x, y|O_1, \dots, O_m) = \frac{P(x, y)}{P(O_1, \dots, O_m)} * \prod_{i=1}^m \frac{P(O_i) * P(x, y|O_i)}{P_i(x, y)}$$

Because the prior  $P(x, y)$  and  $P_i(x, y)$  are assumed to be constant over the area of interest, they cancel out.  $P(O_i)$  and  $P(O_1, \dots, O_m)$  depend only on the observation and are normalizing constants. Therefore, the conditional probability distribution of the target given all observations are defined as follows:

$$P(x, y|O_1, \dots, O_m) = K * \prod_{i=1}^m P(x, y|O_i)$$

Because  $P(x, y|O_i)$  is Gaussian, its conditional probability density function can be defined by:

$$P(\mathbf{x}|\mathbf{O}_1, \dots, \mathbf{O}_m) = K * \exp \left\{ -\frac{1}{2} \sum_{i=1}^m (\mathbf{x} - \mathbf{O}_i)^T \mathbf{C}_i^{-1} (\mathbf{x} - \mathbf{O}_i) \right\}$$

where  $K$  is the scaling factor;  $\mathbf{x}$  is the vector defining target location;  $\mathbf{O}_i$  is the observation of  $\mathbf{x}$  and  $\mathbf{C}_i^{-1}$  is the covariance matrix of  $\mathbf{x}$ . As with the 1-d case, the combined variance will be minimum which will be achieved by differentiating the probability density function given above:

$$\frac{\partial p}{\partial \mathbf{x}} = K * e^{E_m} \left\{ -\frac{1}{2} \frac{\partial E_m}{\partial \mathbf{x}} \right\}$$

where  $E_m$  is defined by:

$$E_m = \sum_{i=1}^m (\mathbf{x} - \mathbf{O}_i)^T \mathbf{C}_i^{-1} (\mathbf{x} - \mathbf{O}_i)$$

$$\frac{\partial E_m}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial E_m}{\partial x_1} \\ \vdots \\ \frac{\partial E_m}{\partial x_n} \end{bmatrix}$$

where  $m$  is the number of observations and  $n$  is the dimensionality of the vector  $\mathbf{x}$  and  $\mathbf{O}$ . Consider the case where there is a single observation,  $\mathbf{O}_1$  then,

$$\begin{aligned} E_1 &= (\mathbf{x} - \mathbf{O}_1)^T \mathbf{C}_1^{-1} (\mathbf{x} - \mathbf{O}_1) \\ &= [x_1 - O_{11} \quad \cdots \quad x_n - O_{1n}] \begin{bmatrix} B_{11} & \cdots & B_{1n} \\ \vdots & & \vdots \\ B_{n1} & \cdots & B_{nn} \end{bmatrix} \begin{bmatrix} x_1 - O_{11} \\ \vdots \\ x_n - O_{1n} \end{bmatrix} \\ &= \left[ \sum_{i=1}^n B_{i1}(x_i - O_{1i}) \quad \cdots \quad \sum_{i=1}^n B_{in}(x_i - O_{1i}) \right] \begin{bmatrix} x_1 - O_{11} \\ \vdots \\ x_n - O_{1n} \end{bmatrix} \\ &= \sum_{j=1}^n (x_j - O_{1j}) \sum_{i=1}^n B_{ij}(x_i - O_{1i}) \end{aligned}$$

Taking the derivative of the above equation with respect to  $x$

$$\frac{\partial E_1}{\partial x_k} = \sum_{i=1}^n B_{ik}(x_i - O_{1i}) + \sum_{j=1}^n B_{kj}(x_j - O_{1j})$$

Because the indices  $i$  and  $j$  both sum to one, the entire equation can be written as follows:

$$\begin{aligned} \frac{\partial E_1}{\partial x_k} &= \sum_{i=1}^n B_{ik}(x_i - O_{1i}) + \sum_{i=1}^n B_{ki}(x_i - O_{1i}) \\ &= 2 \sum_{i=1}^n B_{ki}(x_i - O_{1i}) \\ &= 2\mathbf{C}_1^{-1}(\mathbf{x} - \mathbf{O}_1) \end{aligned}$$

Setting the derivative to zero

$$\begin{aligned} 2\mathbf{C}_1^{-1}(\mathbf{x} - \mathbf{O}_1) &= 0 \\ \mathbf{C}_1^{-1}\mathbf{x} - \mathbf{C}_1^{-1}\mathbf{O}_1 &= 0 \\ \mathbf{x} &= \mathbf{O}_1 \end{aligned}$$

If only one set of observation  $\mathbf{O}_i$  is considered, the best estimation of the target's location  $\mathbf{x}$  will be equal to that observation  $\mathbf{O}_i$  (Orechovesky Jr, 1996). Expanding the estimation by considering  $m$  observations:

$$\begin{aligned}
E_M &= \sum_{m=1}^M \left( \sum_{j=1}^n (x_j - O_{mj}) \sum_{i=1}^n B_{ij} (x_i - O_{mi}) \right) \\
&= \sum_{m=1}^M (\mathbf{x} - \mathbf{O}_m)^T \mathbf{C}_m^{-1} (\mathbf{x} - \mathbf{O}_m)
\end{aligned}$$

Taking the derivative of the above equation with respect to  $\mathbf{x}$

$$\frac{\partial E_n}{\partial \mathbf{x}} = 2 \sum_{i=1}^n \mathbf{c}_i^{-1} (\bar{\mathbf{x}} - \mathbf{O}_i)$$

Setting the derivative to zero

$$\begin{aligned}
\sum_{m=1}^M [\mathbf{c}_m^{-1} (\bar{\mathbf{x}} - \mathbf{O}_m)] &= 0 \\
\left( \sum_{m=1}^M \mathbf{c}_m^{-1} \right) \bar{\mathbf{x}} - \sum_{m=1}^M (\mathbf{C}_m^{-1} \mathbf{O}_m) &= 0 \\
\frac{1}{\left( \sum_{m=1}^M \mathbf{c}_m^{-1} \right)} \left( \sum_{m=1}^M \mathbf{c}_m^{-1} \right) \bar{\mathbf{x}} - \sum_{m=1}^M (\mathbf{C}_m^{-1} \mathbf{O}_m) \frac{1}{\left( \sum_{m=1}^M \mathbf{c}_m^{-1} \right)} &= 0 \\
\bar{\mathbf{x}} - \sum_{m=1}^M \mathbf{C} (\mathbf{C}_m^{-1} \mathbf{O}_m) &= 0
\end{aligned}$$

Finally, the combined mean  $\bar{\mathbf{x}}$  and variance  $\mathbf{C}$  are defined by:

$$\bar{\mathbf{x}} = \mathbf{C} \sum_{m=1}^M (\mathbf{C}_m^{-1} \mathbf{O}_m), \quad \mathbf{C} = \left( \sum_{m=1}^M \mathbf{C}_m^{-1} \right)^{-1}$$

### Example

Consider two bivariate Gaussian distributions,  $\mathbf{X}, \mathbf{Y} \sim N(\boldsymbol{\mu}, \Sigma)$ . The variables of these random vectors are  $\mathbf{X} = [X_1, X_2]$  and  $\mathbf{Y} = [Y_1, Y_2]$ . The parameters of the aforementioned bivariate Gaussian distributions are  $\boldsymbol{\mu}_X = [4, 6]$ ,  $\boldsymbol{\mu}_Y = [3, 8]$  and  $(2 \times 2)$  positive definite covariance matrices  $\Sigma_X, \Sigma_Y$  are:

$$\Sigma_X = \begin{bmatrix} 0.7 & 0.9 \\ 0.9 & 1.9 \end{bmatrix} \quad \Sigma_Y = \begin{bmatrix} 0.5 & 0 \\ 0 & 1.2 \end{bmatrix}$$

Based on the parameters given above, the variables,  $X_1, X_2$  and  $Y_1, Y_2$  were simulated. Let  $\{x_1^k, x_2^l; k, l = 1, \dots, 800\}$  be the simulated samples of  $X_1, X_2$ , and  $\{y_1^k, y_2^l; k, l = 1, \dots, 800\}$  be the simulated samples of  $Y_1, Y_2$ . The histograms of the samples of the simulated variables are given in Figure 3.

The probability level  $(1 - 0.05 = 0.95)$ ,  $\chi_2^2(0.05) = 5.99$  is first selected, as explained in Step 1. This means that 95% probability that the values of the random vectors  $\mathbf{X}$  and  $\mathbf{Y}$  will be inside the ellipses. The eigenvalues (Step 2) of  $\Sigma_X$  and  $\Sigma_Y$  are calculated to be  $\lambda_X = [2.38, 0.22]$  and  $\lambda_Y = [1.2, 0.5]$ , respectively.



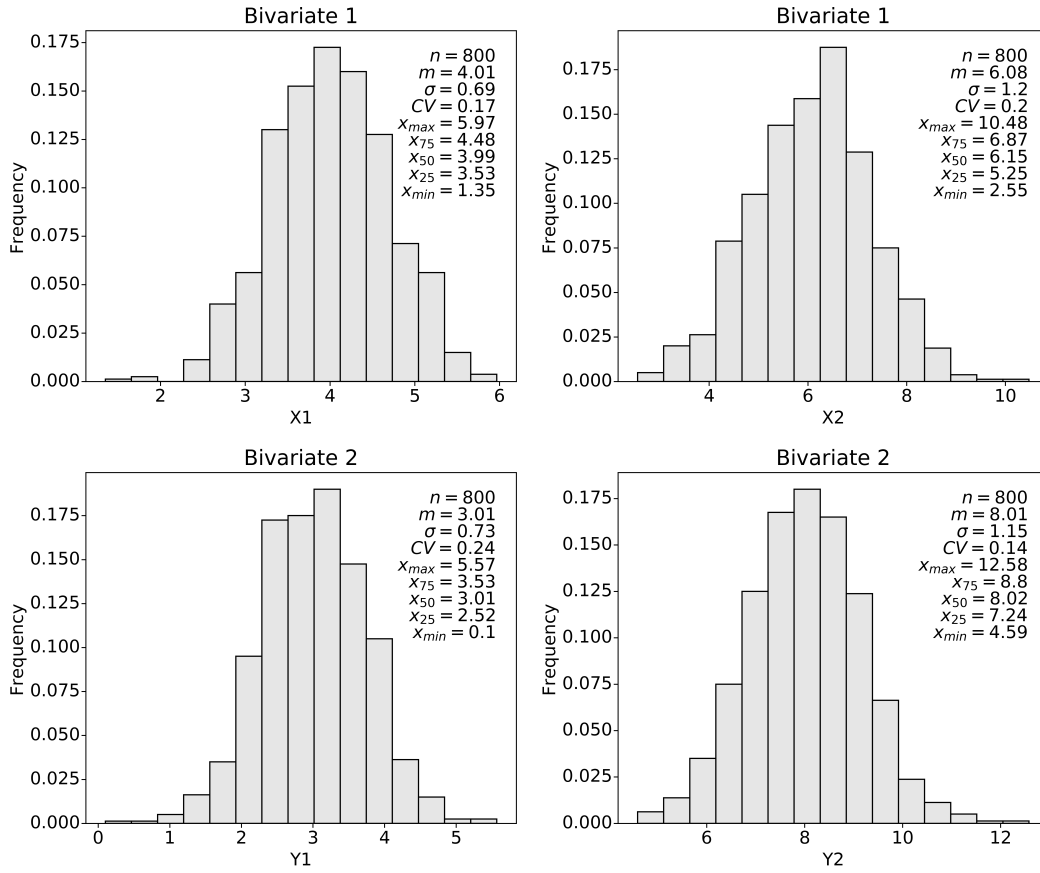


Figure 3: Histograms of the simulated variables

Considering the first bivariate Gaussian distribution defined by the random vector  $\mathbf{X}$ , the variables are positively correlated  $\rho = 0.78$ . Therefore, the axes of the ellipse will point in the directions of the eigenvectors (Step 3). The eigenvectors of  $\Sigma_{\mathbf{X}}$  are calculated to be  $v_1 = [0.88, 0.47]^T$  and  $v_2 = [0.47, -0.88]^T$ . The lengths of the ellipse axes are calculated to be  $(2\sqrt{5.99 \cdot 2.38} = 7.55)$ ,  $(2\sqrt{5.99 \cdot 0.22} = 2.3)$ . The axes of the 95% confidence ellipse of the random vector  $\mathbf{X}$  are then defined by  $\pm\sqrt{5.99 \cdot 2.38} \cdot [0.88, 0.47]^T$  and  $\pm\sqrt{5.99 \cdot 0.22} \cdot [0.47, -0.88]^T$ .

Considering the second bivariate Gaussian distribution defined by the random vector  $\mathbf{Y}$ , the variables are independent. The eigenvalues of  $\Sigma_{\mathbf{Y}}$  are equal to the variances of  $Y_1$  and  $Y_2$ , respectively. As explained in Step 3, the axes of the ellipse are parallel to the original coordinate axes. The lengths of the axes are calculated to be  $(2\sqrt{5.99 \cdot 1.2} = 5.36)$ ,  $(2\sqrt{5.99 \cdot 0.5} = 3.46)$ . The axes of the 95% confidence ellipse of the random vector  $\mathbf{Y}$  are then defined by  $\pm\sqrt{5.99 \cdot 1.2} \cdot [0, 1]^T$  and  $\pm\sqrt{5.99 \cdot 0.5} \cdot [1, 0]^T$ . The 95% confidence ellipses for the given bivariate Gaussian distributions are displayed in Figure 4.

The variance  $C_{\bar{\mathbf{x}}}$  of the combined probability distribution is calculated as follows:

$$C_{\bar{\mathbf{x}}} = \left( \begin{bmatrix} 0.7 & 0.9 \\ 0.9 & 1.9 \end{bmatrix}^{-1} + \begin{bmatrix} 0.5 & 0 \\ 0 & 1.2 \end{bmatrix}^{-1} \right)^{-1}$$

$$C_{\bar{\mathbf{x}}} = \begin{bmatrix} 0.23 & 0.19 \\ 0.19 & 0.61 \end{bmatrix}$$

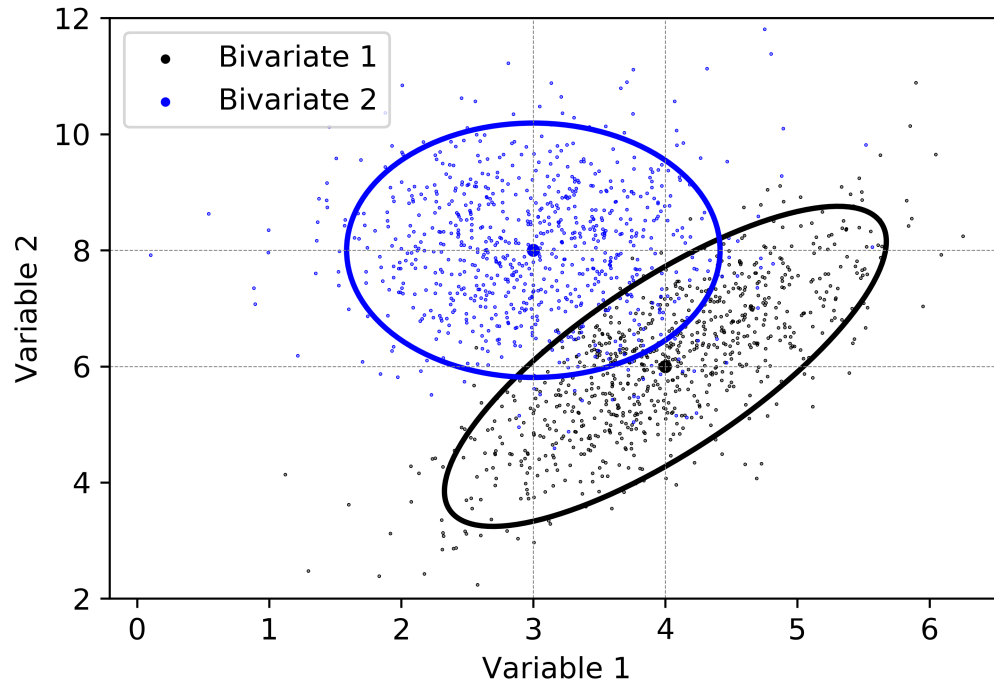


Figure 4: 95% confidence ellipses of the bivariate Gaussian distributions

and the mean vector of the combined distribution  $\bar{\mathbf{x}}$  is then defined by:

$$\bar{\mathbf{x}} = \begin{bmatrix} 0.23 & 0.19 \\ 0.19 & 0.61 \end{bmatrix} \left( \begin{bmatrix} 0.7 & 0.9 \\ 0.9 & 1.9 \end{bmatrix}^{-1} \begin{bmatrix} 4 \\ 6 \end{bmatrix} + \begin{bmatrix} 0.5 & 0 \\ 0 & 1.2 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ 8 \end{bmatrix} \right)$$

$$\bar{\mathbf{x}} = \begin{bmatrix} 3.84 \\ 6.64 \end{bmatrix}$$

Based on the estimated variance  $C_{\bar{\mathbf{x}}}$  and the mean  $\bar{\mathbf{x}}$ , the bivariate Gaussian variables are simulated. 95% confidence ellipse of the combined distribution is given in Figure 5.

## 5 Summary

The error ellipses method combines probability distributions by a linear combination. The combined probability distribution is a convex estimation of the mean of the combined distribution, that is, the estimated mean of the combined distribution always falls between the means of the input probability distributions; the lower variance distribution has the greatest influence. This method is especially useful when there is no clear prior distribution, for example, in the estimation of parameters from diverse data sources. The method should not be used for spatial prediction when a reasonable prior model is available.

## 6 References

Blachman, N. M. (1989). On combining target-location ellipses. *IEEE Transactions on Aerospace and Electronic Systems*, 25(2), 284–287.

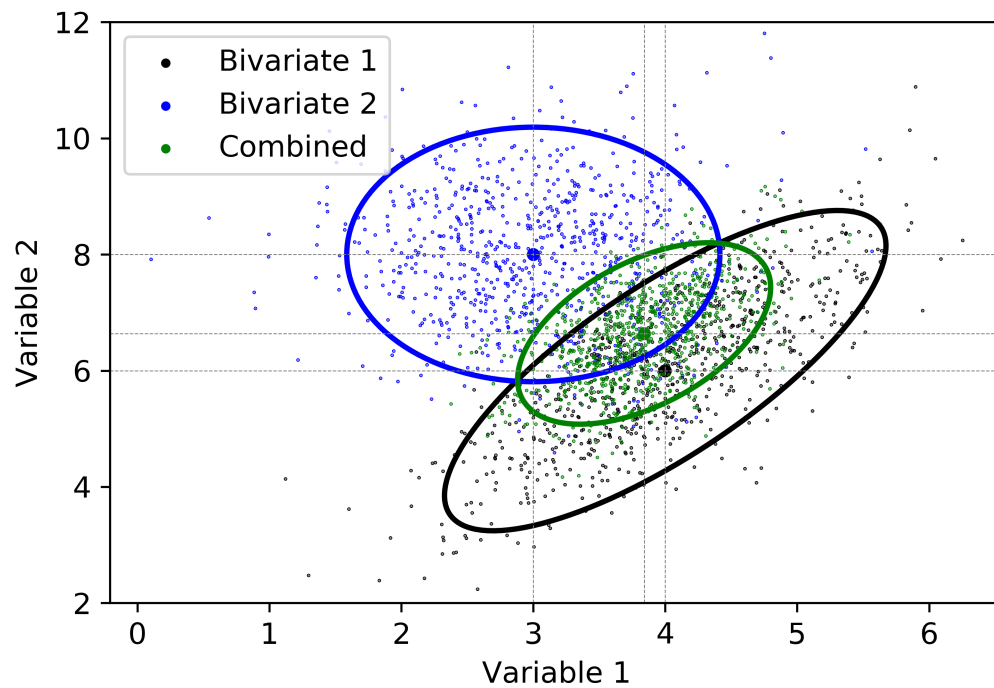


Figure 5: 95% confidence ellipses of the bivariate Gaussian distributions and the combined distribution

- Deutsch, C. V., & Deutsch, J. L. (2018). An Application of Bayes Theorem to Geostatistical Mapping. In J. L. Deutsch (Ed.), *Geostatistics lessons*. Retrieved from <http://geostatisticslessons.com/lessons/bayesmapping>
- Deutsch, C. V., & Journel, A. G. (1998). *GSLIB - geostatistical software library and user's guide*. Oxford University Press.
- Deutsch, J. L., & Deutsch, C. V. (2011). Plotting and checking the bivariate distributions of multiple Gaussian data. *Computers & Geosciences*, 37(10), 1677–1684.
- Doyen, P. M. (1988). Porosity from seismic data: A geostatistical approach. *Geophysics*, 53(10), 1263–1275.
- Doyen, P. M., Boer, L. D. den, & Pillet, W. R. (1996). Seismic porosity mapping in the Ekofisk field using a new form of collocated cokriging. In *SPE annual technical conference and exhibition* (pp. 21–30).
- Johnson, R. A., & Wichern, D. W. (1988). *Applied Multivariate Statistical Analysis*. Prentice-Hall, Inc.
- Orechovesky Jr, J. R. (1996). *Single Source Error Ellipse Combination* (PhD thesis). United States Navy. Retrieved from <https://books.google.ca/books?id=vHD1AAAACAAJ>
- Rezvandehy, M. (2016). *Geostatistical reservoir modeling with parameter uncertainty in presence of limited well data* (PhD thesis). University of Alberta, Edmonton, Canada.
- Xu, W., Tran, T. T., Srivastava, R. M., & Journel, A. G. (1992). Integrating seismic data in reservoir modeling: the collocated cokriging alternative. In *SPE annual technical conference and exhibition* (pp. 833–842).

#### Citation

Erten, O. & Deutsch, C.V. (2020). Combination of Multivariate Gaussian Distributions through Error Ellipses. In J.L. Deutsch (Ed.), *Geostatistics Lessons*. Retrieved from <http://geostatisticslessons.com/lessons/errorellipses>