Computing Multivariate Normal Contour Surfaces

Wagner L. Truppel

January 2001

Given a one-dimensional normal distribution, it's a well known result that approximately 68% of the data lies within one standard deviation from the true mean. This defines a particular 'contour line' composed of just two points, $x_\pm = \mu \pm \sigma$. In this essay, I'll show how to obtain the analogous contour surface in the case of a multivariate normal distribution.

Let

$$p(x) = (2\pi)^{-d/2}\,|\Sigma|^{-1/2}\,\exp\Big[-\frac{1}{2}\,(x-\mu)\cdot\Sigma^{-1}\cdot(x-\mu)\Big]$$

be the (normal) probability density function governing the distribution of a d-dimensional random continuous variable x, with mean μ and covariance matrix Σ.

A contour surface of this function is a $(d-1)$-dimensional surface defined by the equation $p(x) = C =$ constant. The same surface is also obtained by imposing, instead, that $(x-\mu)\cdot\Sigma^{-1}\cdot(x-\mu) = a =$ constant, since the only dependency on x lies in the argument of the exponential function. This surface is, then, a hyper-conic section.[1] As mentioned above, in one dimension the contour line is a set of two points, and the particular choice that contains approximately 68% of the data turns out to be characterized by $a = 1$. In this essay, I'll derive the value of a corresponding to the d-dimensional contour surface which also encompasses approximately 68% of the data. In addition, I'll obtain the equation of the corresponding hyper-elliptical contour surface.

The first order of business is to get rid of the mean, which we can accomplish by transforming from x to $x' = x - \mu$. We next diagonalize the resulting quadratic form $x'\cdot\Sigma^{-1}\cdot x'$ by performing another transformation, $x' = R\cdot z$, where R is an orthogonal

[1] Due to some special properties of the covariance matrix, the contour surface is in fact a hyper-ellipse.

matrix.[2] The original density function is then changed in form (but not in value),

$$\begin{aligned} p(x) \quad \to \quad p'(z) &= (2\pi)^{-d/2} \Big(\prod_{i=1}^{d} \lambda_i \Big)^{-1/2} \exp\Big(-\frac{1}{2} \sum_{i=1}^{d} \lambda_i^{-1} z_i^2 \Big) \\ &= \prod_{i=1}^{d} \Big[(2\pi)^{-1/2} \, \lambda_i^{-1/2} \exp\Big(-\frac{1}{2} \, \lambda_i^{-1} z_i^2 \Big) \Big] , \end{aligned}$$

where $\{\lambda_i | 1 \le i \le d\}$ is the set of eigenvalues of Σ. In order to obtain the desired contour hyper-ellipse $\mathcal{C}$, we now impose the condition

$$\int_{x \subseteq \mathcal{C}} p(x) \, dx = \int_{-1}^{+1} (2\pi)^{-1/2} \, e^{-t^2/2} \, dt \, .$$

The last integral is just the restriction of the left–hand–side integral to the case of one dimension, and is approximately equal to 0.68. Now, because we have diagonalized $p(x)$, the integral in the left–hand–side is easy to perform,[3]

$$\int_{x \subseteq \mathcal{C}} p(x) \, dx = \prod_{i=1}^{d} \Big[\int_{-\alpha_i}^{+\alpha_i} (2\pi)^{-1/2} \, \lambda_i^{-1/2} \exp\Big(-\frac{1}{2} \, \lambda_i^{-1} z_i^2 \Big) \, dz_i \Big] ,$$

where each α_i is chosen to guarantee that x falls within the contour. The transformation $u_i = \lambda_i^{-1/2} \, z_i \, / \sqrt{2}$ allows us to write

$$\int_{x \subseteq \mathcal{C}} p(x) \, dx = \prod_{i=1}^{d} \Big[\frac{2}{\sqrt{\pi}} \int_{0}^{\alpha_i \, \lambda_i^{-1/2} / \sqrt{2}} e^{-u_i^2} \, du_i \Big] .$$

We have not yet specified the contour, however. Recall that the contour is characterized by $(x - \mu) \cdot \Sigma^{-1} \cdot (x - \mu) = a$ for some constant a. In the diagonalized form, this translates into

$$\sum_{i=1}^{d} \lambda_i^{-1} z_i^2 = a \qquad \text{or, equivalently,} \qquad \sum_{i=1}^{d} \Big(\frac{z_i}{\sqrt{a \, \lambda_i}} \Big)^2 = 1 \, .$$

This result shows that the z-variables are constrained to the range $|z_i| \le \sqrt{a \, \lambda_i}$ if x is to lie within the contour. Hence, $\alpha_i = \sqrt{a \, \lambda_i}$, the upper limit of the integral over u_i is independent of i, and we may write

$$\int_{x \subseteq \mathcal{C}} p(x) \, dx = \prod_{i=1}^{d} \Big[\frac{2}{\sqrt{\pi}} \int_{0}^{\sqrt{a/2}} e^{-u_i^2} \, du_i \Big] = \Big[erf\Big(\sqrt{\frac{a}{2}} \Big) \Big]^{d} ,$$

[2]The diagonalization by means of an orthogonal transformation is guaranteed to be possible due to the real and symmetric nature of Σ.

[3]Note that the Jacobian of the transformation from x to z is the matrix R, and since R is orthogonal, its determinant has absolute value 1: $\tilde{R} = R^{-1} \Rightarrow R \cdot \tilde{R} = I \Rightarrow |R|^2 = 1$.

where $erf(z) \equiv \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\, dt$ is the error function.

Since we want a contour surface in any number of dimensions encompassing the same amount of data lying within the contour line in one dimension, we now impose

$$\Big[erf\Big(\sqrt{\frac{a}{2}}\,\Big) \Big]^d = \int_{-1}^{+1} (2\pi)^{-1/2}\, e^{-t^2/2}\, dt = erf(\frac{1}{\sqrt{2}}) \approx 0.68\,,$$

from which it follows

$$a(d) = 2 \left[erf^{-1} \left(\Big[erf\Big(\frac{1}{\sqrt{2}}\Big) \Big]^{1/d} \right) \right]^2 .$$

Particular values of interest are $a \approx 1.85029$ for $d = 2$ and $a \approx 2.42417$ for $d = 3$[4]. Having found a for a given dimension, the equation for the actual contour surface, in the diagonal representation, is given by

$$\sum_{i=1}^{d} (\frac{z_i}{\sqrt{a\,\lambda_i}})^2 = 1\,.$$

Once the z's have been computed, we can revert back to the x representation by virtue of $x = \mu + R \cdot z$. For instance, in two dimensions, we may write (for each distribution)

$$\begin{aligned} z_1 &= \sqrt{a\,\lambda_1}\, \cos\theta\,, \qquad \text{and} \\ z_2 &= \sqrt{a\,\lambda_2}\, \sin\theta \qquad (0 \leq \theta \leq 2\pi)\,, \end{aligned}$$

and it follows that the contour for each distribution can be described by an explicit parametric equation, namely,

$$\begin{cases} x_1 = \mu_1 + R_{11} \sqrt{a\,\lambda_1}\, \cos\theta + R_{12} \sqrt{a\,\lambda_2}\, \sin\theta\,, \\ x_2 = \mu_2 + R_{21} \sqrt{a\,\lambda_1}\, \cos\theta + R_{22} \sqrt{a\,\lambda_2}\, \sin\theta\,. \end{cases}$$

■

[4]If, instead, we require the data to be 'within 2 standard deviations of the mean,' which in one dimension corresponds to encompassing approximately 95% of the data, the corresponding values for a are $a \approx 5.16737$ for $d = 2$ and $a \approx 5.8698$ for $d = 3$.