Note on completing the square

Aasa Feragen

February 13, 2014

Warning: This note is newly written and almost certainly contains typos. If you get stuck at any point, please send a question to the StatML forum or email me at aasa@diku.dk

The process of ”completing the square” is a general trick that allows you to write exponential functions whose exponentials are quadratic, as a Gaussian probability density function with known mean and (co)variance.

Since the main purpose of this note is to understand the underlying concept, let us look at what this means for one-dimensional real variables, and make it an instructional exercise to repeat the procedure for higher-dimensional Gaussian distributions.

1 Completing the square in one dimension

Lemma 1 (Completing the square in one dimension) *If $x \in \mathbb{R}$ is a real variable, any function*

$$f(x) = Ce^{c_1 x^2 + c_2 x + c_3} \tag{2}$$

can be rewritten

$$f(x) = \tilde{C} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \tag{3}$$

for some $\tilde{C} \in \mathbb{R}$, $\mu = \frac{-c_2}{2c_1}$ and $\sigma^2 = -\frac{1}{2c_1}$.

Proof. Verify for yourself that

$$c_1 x^2 + c_2 x + c_3 = c_1(x + \frac{c_2}{2c_1})^2 + (c_3 - \frac{c_2^2}{4c_1}) \tag{4}$$

by multiplying out the right hand side above.

Substituting (4) in the exponential (2), we see that

$$\begin{aligned} f(x) &= Ce^{c_1 x^2 + c_2 x + c_3} \\ &= Ce^{(c_3 - \frac{c_2^2}{4c_1}) + c_1(x + \frac{c_2}{2c_1})^2} \\ &= \underbrace{Ce^{c_3 - \frac{c_2^2}{4c_1}}}_{const} e^{c_1(x + \frac{c_2}{2c_1})^2} \\ &= \tilde{C} e^{c_1(x - \frac{-c_2}{2c_1})^2} \\ &= \tilde{C} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \end{aligned}$$

if

$$\tilde{C} = Ce^{c_3 - \frac{c_2^2}{4c_1}}, \quad \mu = \frac{-c_2}{2c_1}, \quad -\frac{1}{2\sigma^2} = c_1,$$

where the last equality gives $\sigma = \sqrt{-\frac{1}{2c_1}}$[1]. It follows that $f(x)$ can be written as in Eq. (3) with $\mu = \frac{-c_2}{2c_1}$ and $\sigma^2 = -\frac{1}{2c_1}$. □

Remark 5 Note that f is not necessarily a probability density function – you need to normalize it so that $\int_{-\infty}^{\infty} f(x)dx = 1$.

2 Example: Conditional of Gaussian

The "completing the square" trick allows you to find the mean and variance of the corresponding Gaussian distribution for any function of the form (2)! We are going to use this for computing the conditional of a Gaussian distribution in 2 variables x_a and x_b.

A formula for the conditional $p(x_a|x_b)$ can be derived from the product rule:

$$p(x_a, x_b) = p(x_a|x_b)p(x_b),$$

which gives

$$p(x_a|x_b) = \frac{p(x_a, x_b)}{p(x_b)}. \tag{6}$$

Thus, we can derive a formula for $p(x_a|x_b)$ directly from the formula for $p(x_a, x_b)$ by fixing x_b to its fixed value and normalizing the resulting exponential function so that it integrates to 1. We could perform this normalization explicitly, but it is easier to do this using the trick of *completing the square* as defined above.

From now on, we are going to consider x_b to be a constant, and x_a to be a variable. Since the joint probability distribution $p(x_a, x_b)$ is a Gaussian $\mathcal{N}(\mu, \mathbf{\Sigma})$, its restriction to fixed x_b is given by the formula

$$g(x_a) = Ce^{\frac{1}{2}(\mathbf{x}-\mu)^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)}, \tag{7}$$

where

$$\mathbf{x} = \begin{pmatrix} x_a \\ x_b \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}, \quad \mathbf{\Sigma} = \begin{pmatrix} \sigma_{aa} & \sigma_{ab} \\ \sigma_{ba} & \sigma_{bb} \end{pmatrix}, \quad \mathbf{\Lambda} = \mathbf{\Sigma}^{-1} = \begin{pmatrix} \lambda_{aa} & \lambda_{ab} \\ \lambda_{ba} & \lambda_{bb} \end{pmatrix}, \tag{8}$$

and C is a normalization factor given by

$$C = \frac{1}{2\pi\sqrt{|\mathbf{\Sigma}|}}.$$

The function $g(x_a)$ is an unnofmalized version of probability density function $p(x_a|x_b)$ which we are looking for.

[1] That is, for σ to be a real standard deviation, we need $c_1 < 0$.

Using the decompositions in (8) we can expand (7) as a function of x_b so that we can integrate it. All our analysis will be done in the exponent of the exponential function, so for the sake of doing algebra, let's just look at the exponent of (7). We can work out:

$$\begin{aligned} &-\tfrac{1}{2}(\mathbf{x}-\mu)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mu) \\ &= -\tfrac{1}{2}\begin{bmatrix} (x_a-\mu_a) & (x_b-\mu_b) \end{bmatrix}\begin{bmatrix} \lambda_{aa} & \lambda_{ab} \\ \lambda_{ba} & \lambda_{bb} \end{bmatrix}\begin{bmatrix} x_a-\mu_a \\ x_b-\mu_b \end{bmatrix} \end{aligned} \tag{9}$$

Performing the matrix product step by step, we get:

$$\begin{aligned} &-\tfrac{1}{2}\begin{bmatrix} (x_a-\mu_a) & (x_b-\mu_b) \end{bmatrix}\begin{bmatrix} \lambda_{aa}(x_a-\mu_a)+\lambda_{ab}(x_b-\mu_b) \\ \lambda_{ba}(x_a-\mu_a)+\lambda_{bb}(x_b-\mu_b) \end{bmatrix} \\ &\quad = -\tfrac{1}{2}(x_a-\mu_a)\left(\lambda_{aa}(x_a-\mu_a)+\lambda_{ab}(x_b-\mu_b)\right)+ \\ &\quad\quad (x_b-\mu_b)\left(\lambda_{ba}(x_a-\mu_a)+\lambda_{bb}(x_b-\mu_b)\right) \end{aligned} \tag{10}$$

which we can expand out as

$$\begin{aligned} &-\tfrac{1}{2}\left((x_a-\mu_a)\lambda_{aa}(x_a-\mu_a)+(x_a-\mu_a)\lambda_{ab}(x_b-\mu_b)\right. \\ &\quad \left.+(x_b-\mu_b)\lambda_{ba}(x_a-\mu_a)+(x_b-\mu_b)\lambda_{bb}(x_b-\mu_b)\right), \end{aligned} \tag{11}$$

which we can rearrange as

$$\underbrace{[-\frac{1}{2}\lambda_{aa}]}_{c_1}x_a^2+\underbrace{[\lambda_{aa}\mu_a-\lambda_{ab}x_b+\lambda_{ab}\mu_b]}_{c_2}x_a+\underbrace{\text{terms that do not depend on } x_a}_{c_3} \tag{12}$$

Remark 13 Verify for yourself that (11) holds, and compare this equation with the higher dimension equation (2.70) in CB p 86!

Referring back to Lemma 1, this means that we can write

$$g(x_a)=Ce^{c_1x_a^2+c_2x_a+c_3}=\tilde{C}e^{-\frac{1}{2\sigma_{a|b}^2}(x_a-\mu_{a|b})^2}, \tag{14}$$

where

$$\mu_{a|b}=\frac{-c_2}{2c_1}=\frac{-[\lambda_{aa}\mu_a-\lambda_{ab}x_b+\lambda_{ab}\mu_b]}{2[-\frac{1}{2}\lambda_{aa}]}=\mu_a-\frac{\lambda_{ab}}{\lambda_{aa}}(x_b-\mu_b),$$

and

$$\sigma_{a|b}^2=-\frac{1}{2c_1}=-\frac{1}{2[-\frac{1}{2}\lambda_{aa}]}=\frac{1}{\lambda_{aa}}$$

Remark 15 Again, compare with the corresponding equations (2.73) and (2.74) in CB p. 86-87.

Based on the above we can conclude that the conditional probability distribution $p(x_a|x_b)$ is a 1-dimensional Gaussian distribution on the form

$$\tilde{C}e^{-\frac{1}{2\sigma_{a|b}^2}(x_a-\mu_{a|b})^2},$$

but we are missing the normalization factor $\tilde{C}$. Recall, however, that the formula for the 1-dimensional Gaussian with mean $\mu_{a|b}$ and variance $\sigma_{a|b}^2$ is

$$\mathcal{N}(x_a|\mu_{a|b}, \sigma_{a|b}) = \frac{1}{\sqrt{2\pi\sigma_{a|b}^2}} e^{-\frac{1}{2\sigma_{a|b}^2}(x_a - \mu_{a|b})^2},$$

so in particular

$$\tilde{C} = \frac{1}{\sqrt{2\pi\sigma_{a|b}^2}},$$

and $p(x_a|x_b) = \mathcal{N}(x_a|\mu_{a|b}, \sigma_{a|b})$.

3 Completing the square in more dimensions

Completing the square in higher dimensions follows exactly the same procedure. If the matrix/vector notation confuses you, write out the vectors like we did with the 2-dimensional vectors above and go through the algebraic computations, step by step.