On the corner-point parametrization

Rodrigo Labouriau *

May, 2014

Where is the centre of the solar system, at the sun or at the earth? Placing the centre at the earth yields very complex descriptions of the movements of the stars and planets in the sky, but in the sixteenth century you would had avoided to be burned. Placing the sun at the centre gives elegant descriptions of the movements of stars and planets, but in the sixteenth century the inquisition ... But, who cares on the location of the centre of the solar system nowadays that any modest computer would move back and through from one to another representation and calculate the positions and the movements of planets in a tiny fraction of second?

This text will explain how the corner point parametrisation works in a one way and a two-ways classification model.

1 The case with one factor (one-way)

Suppose that we have a classification variable, say T, which classifies n observations into three (mutually exclusive) categories, t_1, t_2 and t_3. That is, each of the observations belongs to one (and only one) of the three categories. Moreover, none of the three categories is empty, *i.e.* there is at least one observation classified in each of the categories. A certain quantity, Y, is measured in all the observations and is modelled by a statistical model.

*Department of Molecular Biology and Genetics, Aarhus University.

The reader can keep in mind the situation where an experiment with three treatments is performed. For simplicity we consider the case where Y is normally distributed and have a variance equal to 1. That is, $Y \sim N(\mu, 1)$. Furthermore, the model specifies that the expectations of each observation depends only on the category at which the observation is classified. More precisely, the data consists of n pairs $(Y_1, T_1), \ldots, (Y_n, T_n)$, each pair representing the observed quantity Y and the classification T we have on each observation. The data is modelled as being normally distributed according to the following rule: For i from 1 to n

$$Y_i \sim N(\alpha_{T_i}, 1)\,. \tag{1}$$

Here α_1, α_2 and α_3 are values, termed *parameters*, that specify the distributions of the observations. This model is a *one-way statistical model* (with 3 levels) and the classification variable T is a *factor* (with 3 levels). The discussion that follows is valid for other number of classification categories (not necessarily 3) and for other types of models (*eg* a binomial model indexed by the probabilities or the logarithm of the odds).

The statistical model given by (1) is constructed in such a way that we can attribute a probability law for each of the n observations, provided we know the values α_1, α_2 and α_3. To do so we just use the rule:

"For $i = 1, \ldots, n$, the response Y_i of the ith observation follows the normal distribution with expectation α_{T_i} and variance 1." (2)

We call the values α_1, α_2 and α_3 *parameters* and we say that they *parametrize* the statistical model, since they determine completely the distribution of each of the observations (through the rule given by (2)).

The way we parametrize this simple model is arbitrary. For example, we could have used a new set of parameters, $\tilde{\alpha}_1$, $\tilde{\alpha}_2$ and $\tilde{\alpha}_3$, defined by subtracting a fixed value α_* to the original parameters α_1, α_2 and α_3 and at the same time replaced the attribution rule (2) by the rule:

"For $i = 1, \ldots, n$, the response Y_i of the ith observation follows the normal distribution with expectation $\tilde{\alpha}_{T_i} + \alpha_*$ and variance 1." (3)

Note that we specify exactly the same statistical model if we use the parameters α_1, α_2 and α_3 together with the attribution rule (2) or the parameters $\tilde{\alpha}_1$, $\tilde{\alpha}_2$ and $\tilde{\alpha}_3$ together with the attribution rule (3), since the same probability law is specified for each observation. When using the parametrization given by $\tilde{\alpha}_1$, $\tilde{\alpha}_2$ and $\tilde{\alpha}_3$ and the rule (3) we just changed the convention of how we label the probability laws (by changing from α_1, α_2 and α_3 to $\tilde{\alpha}_1$, $\tilde{\alpha}_2$ and $\tilde{\alpha}_3$) and making the corresponding change in the attribution rule so that the same probability law is attributed to each observation. In the first parametrization we used the expectation of each category to parametrize the model and in the second parametrization we used the expectations minus the pre-fixed value α_*. Note that the parametrization defined using $\tilde{\alpha}_1$, $\tilde{\alpha}_2$ and $\tilde{\alpha}_3$ coincides with the parametrization given by α_1, α_2 and α_3 when α_* is equal to zero.

The *corner point parametrization* [1] is a particular case of the parametrization given by $\tilde{\alpha}_1$, $\tilde{\alpha}_2$ and $\tilde{\alpha}_3$ (together with the attribution rule (3)) obtained when the reference level α_* is equal to the value of the parameter for one of the levels of the classification variable T. We will use here the first level, *i.e.* we will set α_* equal to α_1, to define the corner-point parametrization. [2] Clearly, $\tilde{\alpha}_1 = \alpha_1 - \alpha_* = \alpha_1 - \alpha_1 = 0$. So, in the corner-point parametrization we use three parameters (values), $\alpha_* = \alpha_1$, $\tilde{\alpha}_2$ and $\tilde{\alpha}_3$, to describe the model.

We can convert the original parametrization to the corner-point parametrization by subtracting to all the parameters the parameter associated with the first level of the classification variable. Conversely, we can obtain the original parametrization by summing the reference parameter to the parameters associated to each level of the classification variable.

[1] I will justify this name latter.

[2] We could have used other levels as well, yielding the same kind of parametrization.

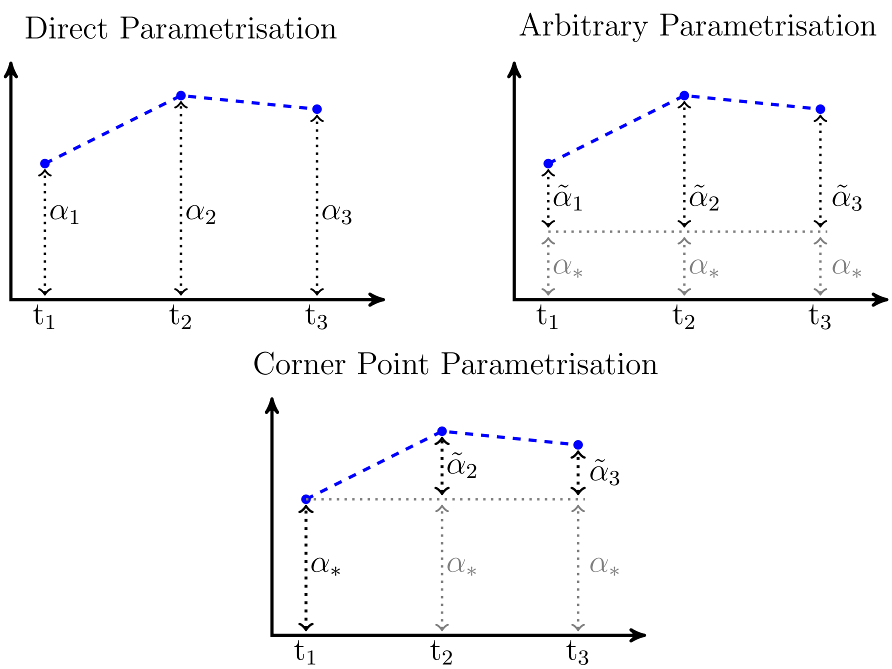

Figure 1: Representation of the direct parametrisation, a parametrisation with an arbitrary reference α_* and the corner point parametrisation for a one-way classification model. The vertical axis represents the expected values of the response. The three levels of the factor T are represented in the horizontal axis (the dashed blue line is just a convenient visual aid).

2 The case with two factors (two-ways)

Suppose now that we have two classification variables, T with the three categories t_1, t_2 and t_3, and S with two categories s_1 and s_2. As before, all the observations are classified according to the classification variables S and T and we assume for simplicity that all the combinations of the two classifications are present in the data. T and S are used in a statistical model as factors (i.e. explanatory variables taking finite number of values) and we observe for each of the, say n, observations the triples $(T_1, S_1, Y_1), \dots (T_n, S_n, Y_n)$. The model states that the distribution of the response variable Y depends only on the categories of the two classification variables T and S. Continuing with our (hypothetical) example using the normal distribution, the data is modelled as being normally distributed according to the following rule: For i from 1 to n

$$Y_i \sim N(\alpha_{S_i,T_i}, 1)\,. \tag{4}$$

Here the expectations are specified $\alpha_{1,1}, \dots, \alpha_{2,3}$ by two indices, the first giving the classification of the observation according to the classification variable S and the second subindex determining the classification according to the factor T. We use then 6 parameters to describe the probability laws of the observations. This six parameters can be displayed in a table below

	t_1	t_2	t_3
s_1	$\alpha_{1,1}$	$\alpha_{1,2}$	$\alpha_{1,3}$
s_2	$\alpha_{2,1}$	$\alpha_{2,2}$	$\alpha_{2,3}$

where the two rows represent the two categories of the classification variable S and the three columns represent the categories of the classification variable T.

The statistical model given by (4) is constructed in such a way that we can attribute a probability law for each of the n observations, provided we know the six parameters $\alpha_{1,1}, \ldots, \alpha_{2,3}$ and we use the attribution rule

"The response Y_i of the ith observation follows the normal distribution with expectation α_{S_i,T_i} and variance 1 (for $i = 1, \ldots, n$)". (5)

Again, the way we parametrize this simple model is arbitrary. We could have used as well a new set of parameters,

	t_1	t_2	t_3
s_1	$\tilde{\alpha}_{1,1} = \alpha_{1,1} - \alpha_*$	$\tilde{\alpha}_{1,2} = \alpha_{1,2} - \alpha_*$	$\tilde{\alpha}_{1,3} = \alpha_{1,3} - \alpha_*$
s_2	$\tilde{\alpha}_{2,1} = \alpha_{2,1} - \alpha_*$	$\tilde{\alpha}_{2,2} = \alpha_{2,2} - \alpha_*$	$\tilde{\alpha}_{2,3} = \alpha_{2,3} - \alpha_*$

where α_* is a given reference value, and the new attribution rule

"The response Y_i of the ith observation follows the normal distribution with expectation $\alpha_{S_i,T_i} + \alpha_*$ and variance 1 (for $i = 1, \ldots, n$)". (6)

The corner-point parametrization is obtained by using the parametrization above and setting the reference value α_* equal to α_{11}, i.e. the parameter associated with the first category (or level) of **both** factors. The table of parameters of the corner-point parametrization becomes:

	t_1	t_2	t_3
s_1	0	$\tilde{\alpha}_{1,2} = \alpha_{1,2} - \alpha_{1,1}$	$\tilde{\alpha}_{1,3} = \alpha_{1,3} - \alpha_{1,1}$
s_2	$\tilde{\alpha}_{2,1} = \alpha_{2,1} - \alpha_{1,1}$	$\tilde{\alpha}_{2,2} = \alpha_{2,2} - \alpha_{1,1}$	$\tilde{\alpha}_{2,3} = \alpha_{2,3} - \alpha_{1,1}$

When using the corner-point parametrization we use six parameters to describe the model, namely the reference parameter (sometimes termed "intercept") α_* and $\tilde{\alpha}_{1,2}, \tilde{\alpha}_{1,3}, \tilde{\alpha}_{2,1}, \tilde{\alpha}_{2,2}$ and $\tilde{\alpha}_{2,3}$. The name of the corner-point parametrization comes from the fact that the upper left corner is per construction zero, and therefore the reference.

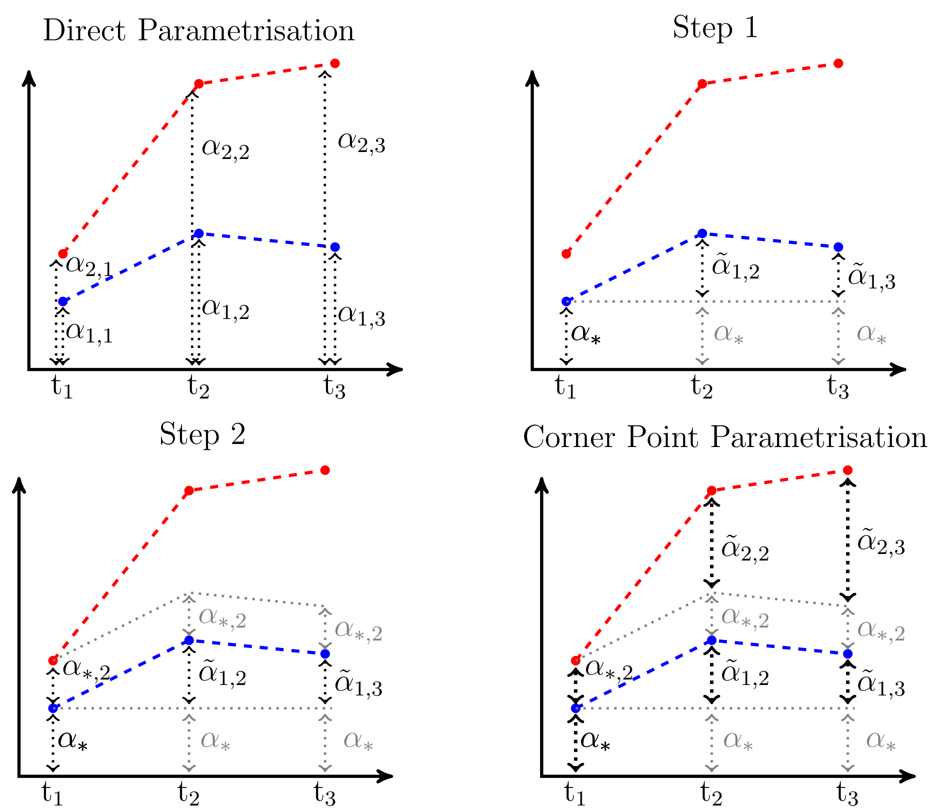

Figure 2: Representation of the direct and the corner point parametrisation for a two-ways classification model. The vertical axis represents the expected values of the response. The three levels of the factor T are represented in the horizontal axis, and the two levels of the factor S are represented by colors: blue for the first level and red for the second level of S (the dashed blue and red lines are just convenient visual aides). The corner point parametrisation is constructed here in three steps. In step 1 a one-way corner point parametrisation is constructed for the first level of S (represented in blue in the upper right graph); in step 2 the profile of first level of S is shifted (parallel transported) in such a way that its first level of T coincides with $\alpha_{2,1}$; in the last step the gaps in the parametrisation are completed by adding the parameters $\tilde{\alpha}_{2,2}$ and $\tilde{\alpha}_{2,3}$ representing the differences between the two profiles (blue and red).