A Bayesian Test for Comparing Classifier Errors

Emanuele Olivetti*†, Dirk B. Walther‡
*NeuroInformatics Laboratory (NILab), Bruno Kessler Foundation, Trento, Italy, olivetti@fbk.eu
†Centro Interdipartimentale Mente e Cervello (CIMeC), University of Trento, Italy
†Department of Psychology, University of Toronto, Canada, bernhardt-walther@psych.utoronto.ca

***Abstract*—Multi-class classification algorithms have become an important tool for the analysis of neuroimaging data. Classification errors contain potentially important information that often goes unreported. It is therefore desirable to quantitatively compare patterns of errors between different experimental conditions. Here we present a Bayesian test that is based on comparing evidence in favor of two competing hypotheses, one stating dependence and one stating independence of two given error patterns. We derive analytical solutions for the likelihoods of both hypotheses. We compare the results from our new test with two other methods of comparing error patterns using data from an fMRI experiment and we substantiate reasons for adopting our proposal and for future work.**

***Index Terms*—fMRI decoding; error comparison; confusion matrix; Bayesian test**

I. Introduction

Successful prediction of a stimulus or behavior from fMRI data in a particular region of interest (ROI) indicates involvement of the ROI in the processing of the stimulus or in the generation of the behavior. However, the specific pattern of prediction errors can yield additional insights. For instance, high confusability of two stimulus exemplars could point to the perceptual features that underlie neural encoding of the stimuli. Rather than auguring such relations from complex error patterns, we wish to find a way to quantify their similarity.

Classifier predictions are usually tabled in a confusion matrix, with correct predictions on the diagonal and prediction errors in off-diagonal cells. Comparison of the error patterns between two experimental conditions can be achieved, for instance, by correlating the off-diagonal elements of two confusion matrices [1] or by computing their mutual information [2]. However, interpretation of these measures and assessing their statistical significance can be problematic.

Here we aim to ameliorate this problem with a statistical method for testing whether two confusion matrices, C_1 and C_2, share the same pattern of errors, i.e. the same pattern in the off-diagonal elements. Specifically, we propose a Bayesian test that is based on comparing the following two hypotheses:

- H_1: the off-diagonal values of C_1 and C_2 are drawn from the *same* distribution.
- H_2: the off-diagonal values of C_1 and C_2 are drawn from two *different* distributions.

We assume that in neuroimaging experiments the stimulation protocol is usually pre-designed. For this reason the number of stimuli per condition is known in advance. In terms of confusion matrices, assuming that true class labels are on the rows and predictions on the columns, the row-marginals of the confusion matrices are known and fixed. Examples of two confusion matrices C_1 and C_2 are shown in Figure 1A. In Figure 1B, the related matrices E_1 and E_2, made by removing the diagonal values, are shown. Notice that the diagonal elements of C_1 and C_2 are not considered here, because here we are only concerned with classifier *errors*.

A

True \ Predicted	**1**	**2**	**3**	**4**	Tot
1	5	5	3	3	16
2	1	13	0	2	16
3	2	0	13	1	16
4	4	2	4	6	16

True \ Predicted	**1**	**2**	**3**	**4**	Tot
1	6	5	3	2	16
2	5	7	2	3	16
3	3	2	7	4	16
4	2	4	4	6	16

B

True \ Errors	1	2	3	Tot
1	5	3	3	11
2	1	0	2	3
3	2	0	1	3
4	4	2	4	10

True \ Errors	1	2	3	Tot
1	5	3	2	10
2	5	2	3	9
3	3	2	4	9
4	2	4	4	10

Fig. 1. A. Example of two 4×4 confusion matrices, C_1 on the left, C_2 on the right. B. Corresponding off-diagonal elements, i.e. errors, together with the row marginals. The new matrices are called E_1 and E_2, respectively.

II. Methods

In this section we describe two generative models for the pair (E_1, E_2), associated with H_1 and H_2. Then we use these models to compute the amount of evidence the observed (E_1, E_2) carry in favor of each of the two hypotheses, both in terms of posterior probabilities, i.e. $p(H_1|E_1, E_2)$ and $p(H_2|E_1, E_2)$, and Bayes factor (see Section II-C).

A. The Generative Models

A generative model for the off-diagonal values of a confusion matrix with fixed row-marginals, can be defined as multinomial for each row. In the example of Figure 1B, the off-diagonal elements E_1 of the first row of C_1 are $(5, 3, 3)$, and the row-marginal of E_1 is 11. Then the multinomial generative model for those off-diagonal elements is $Mul(n = n_{1*}, \theta^1_{11}, \theta^1_{12}, \theta^1_{13})$, where $n_{1*} = 11$ is the first row marginal of E_1, and $\theta^k_{ij} = P(\text{error} = j|\text{True} = i)$ is the conditional probability of each element of matrix E_k. The same kind of generative model with, in general, different parameters, describes each row of each of the two matrices. For this example this means that four multinomial distributions, each generating three values, are defined for each confusion matrix. See Figure 2 for an illustration of the parameters $\{\theta^k_{ij}\}_{ijk}$.

A missing element in the previous model is the definition of the parameters of the models, i.e. $\{\theta^k_{ij}\}_{ijk}$. As is typical

978-1-4673-7145-2/15 $31.00 © 2015 IEEE
DOI 10.1109/PRNI.2015.11

		Cond.Prob.			
		1	2	3	Tot
True	**1**	θ^1_{11}	θ^1_{12}	θ^1_{13}	1
	2	θ^1_{21}	θ^1_{22}	θ^1_{23}	1
	3	θ^1_{31}	θ^1_{32}	θ^1_{33}	1
	4	θ^1_{41}	θ^1_{42}	θ^1_{43}	1

		Cond.Prob.			
		1	2	3	Tot
True	**1**	θ^2_{11}	θ^2_{12}	θ^2_{13}	1
	2	θ^2_{21}	θ^2_{22}	θ^2_{23}	1
	3	θ^2_{31}	θ^2_{32}	θ^2_{33}	1
	4	θ^2_{41}	θ^2_{42}	θ^2_{43}	1

Fig. 2. Conditional probabilities for E_1 and E_2, where $\theta^k_{ij} = p(\text{error} = j|\text{True} = i)$ for classifier k.

in Bayesian models, parameters have their own distributions. Next, we define two generative models for the two hypotheses H_1 and H_2 by defining two different ways to generate $\{\theta^k_{ij}\}_{ijk}$.

Under H_1 we assume that E_1 and E_2 are generated by the exact same mechanism. This means that $\{\theta^1_{ij}\}_{ij} = \{\theta^2_{ij}\}_{ij}$, i.e. we need to generate only $\{\theta^1_{ij}\}_{ij}$ and then $\{\theta^2_{ij}\}_{ij}$ will be identical. Conversely, under H_2, $\{\theta^1_{ij}\}_{ij}$ and $\{\theta^2_{ij}\}_{ij}$ will be generated independently.

A common generative model for θs of multinomial distributions is the Dirichlet distribution $Dir(\alpha_1, \alpha_2, \ldots)$, where $\{\alpha_i\}_i$ are the parameters of the distribution. In case there is no prior information on how the θs of the multinomials should be distributed, a non-informative (flat) Dirichlet distribution is assumed, i.e. $\alpha_i = 1$, $\forall i$. This prior is meaningful for our case since, in general, we have no prior knowledge about the pattern of the off-diagonal values of the confusion matrix. Nevertheless, it is straightforward to extend the derivations in this paper to generic values of α_i.

Summing up, our generative model of the off-diagonal elements of the confusion matrices starts by sampling a set of θs from $Dir(1, 1, \ldots)$ for each row of E_1 and E_2. These $\theta^1_{12}, \theta^1_{13}, \theta^1_{14}$ are then used to instantiate the multinomial distribution of the off-diagonal values. That is, $e^1_{12}, e^1_{13}, e^1_{14}$ are sampled from $Mul(n_{1*}, \theta^1_{12}, \theta^1_{13}, \theta^1_{14})$. In case of H_1 the off-diagonal values of E_2 will be sampled with the same θs as E_1, whereas in case of H_2, each set of values will be sampled from a different draw of θs.

B. Likelihoods $p(E_1, E_2|H_i)$

With the help of the generative model defined so far, we can now derive the likelihoods $p(E_1, E_2|H_1)$ and $p(E_1, E_2|H_2)$, so that we can compute the posterior probabilities, i.e. $p(H_1|E_1, E_2)$, $p(H_2|E_1, E_2)$, and the Bayes factor later.

We recall that Bayesian likelihoods, also called *integrated likelihoods*, are defined by integrating over the parameter space of the generative model according to the parameter distributions. In case of H_1 and H_2, the corresponding generic definitions of the likelihoods are:

$$\begin{aligned}\mathcal{L}_1 &= p(E_1, E_2|H_1) = \int p(E_1, E_2|\boldsymbol{\theta}) p(\boldsymbol{\theta}|H_1) d\boldsymbol{\theta} = \\ &= \int p(E_1|\boldsymbol{\theta}) p(E_2|\boldsymbol{\theta}) p(\boldsymbol{\theta}|H_1) d\boldsymbol{\theta} \end{aligned} \quad (1)$$

$$\begin{aligned}\mathcal{L}_2 &= p(E_1, E_2|H_2) = p(E_1|H_2) p(E_2|H_2) = \\ &= \int p(E_1|\boldsymbol{\theta_1}) p(\boldsymbol{\theta_1}|H_2) d\boldsymbol{\theta_1} \int p(E_2|\boldsymbol{\theta_2}) p(\boldsymbol{\theta_2}|H_2) d\boldsymbol{\theta_2}\end{aligned} \quad (2)$$

where, in case of H_1, $\boldsymbol{\theta}$ is $\{\theta_{ij}\}_{ij}$ (we dropped the index k because under H_1 both E_1 and E_2 share the same θs); in case of H_2, $\boldsymbol{\theta_k}$ is $\{\theta^k_{ij}\}_{ij}$.

Inserting the generative models defined before (see Section II-A) we get:

$$\begin{aligned}\mathcal{L}_1 &= p(C1, C2|H_1) = \\ &= \int \prod_i Mul(e^1_{i1}, \ldots | n^1_{i*}, \theta_{i1}, \ldots) \\ &\quad Mul(e^2_{i1}, \ldots | n^2_{i*}, \theta_{i1}, \ldots) Dir(\theta_{i1}, \ldots | 1, \ldots) d\theta_{i1}, \ldots = \\ &= \prod_i \frac{\frac{n^1_{i*}!}{e^1_{i1}!\ldots} \frac{n^2_{i*}!}{e^2_{i1}!\ldots}}{\frac{(n^1_{i*}+n^2_{i*})!}{(e^1_{i1}+e^2_{i1})!\ldots}} Pólya((e^1_{i1} + e^2_{i1}), \ldots | 1, \ldots). \end{aligned} \quad (3)$$

and

$$\begin{aligned}\mathcal{L}_2 &= p(C1, C2|H_2) = \\ &= \int \prod_i Mul(e^1_{i1}, \ldots | n^1_{i*}, \theta^1_{i1}, \ldots) \\ &\quad Dir(\theta^1_{i1}, \ldots | 1, \ldots) d\theta^1_{i1}, \ldots \\ &\quad \int \prod_i Mul(e^2_{i1}, \ldots | n^2_{i*}, \theta^2_{i1}, \ldots) \\ &\quad Dir(\theta^2_{i1}, \ldots | 1, \ldots) d\theta^2_{i1}, \ldots \\ &= \prod_i Pólya(e^1_{i1}, \ldots | 1, \ldots) Pólya(e^2_{i1}, \ldots | 1, \ldots) \end{aligned} \quad (4)$$

where $Pólya()$ is the Dirichlet compound multinomial (DCM) distribution, also known as multivariate Pólya distribution. See the Appendix for a detailed derivation of $\mathcal{L}_1$ and the definition of the DCM distribution.

C. The Bayesian Test for Confusion Errors

The standard Bayesian way to compare two hypotheses in the light of the data is either to compute the *posterior odds ratio* or the *Bayes factor*, see [3]. While the posterior odds ratio is the ratio of the posterior probabilities of the two hypotheses, which combines the prior knowledge and the evidence of the data, the Bayes factor is the ratio of the likelihoods, which shows the amount of *evidence* the data carries for each hypothesis.

The posterior odds ratio is defined as:

$$odds_{12} = \frac{p(H_1|data)}{p(H_2|data)} \quad (5)$$

and the Bayes factor as:

$$BF_{12} = \frac{p(data|H_1)}{p(data|H_2)} = \frac{\mathcal{L}_1}{\mathcal{L}_2}. \quad (6)$$

The relation between the two quantities is defined by Bayes' theorem:

$$p(H_i|data) = \frac{p(data|H_i)p(H_1)}{p(data)} = \frac{p(data|H_i)p(H_1)}{p(data|H_1) + p(data|H_2)} \quad (7)$$

which gives

$$odds_{12} = \frac{p(H_1|data)}{p(H_2|data)} = \frac{p(data|H_1)}{p(data|H_2)} \frac{p(H_2)}{p(H_1)} = BF_{12} \frac{p(H_2)}{p(H_1)}. \quad (8)$$

$\mathbf{BF_{12}}$	$\mathbf{\log_{10} BF_{12}}$	**Evidence that supports** H_1
< 1	< 0	Negative (supports H_2)
1 to 3	0 to 0.477	Barely worth mentioning
3 to 10	0.477 to 1	Substantial
10 to 30	1 to 1.477	Strong
30 to 100	1.477 to 2	Very strong
> 100	> 2	Decisive

Fig. 3. Guidelines for the interpretation of the value of the Bayes factor (BF_{12}), from [3].

In case of equal prior probability for each hypothesis, i.e. $p(H_1) = p(H_2)$, then $odds_{12} = BF_{12}$.

Here we adopt the Bayes factor (BF). Standard guidelines for the interpretation of BF according to [3] are shown in Figure 3. BF for the proposed Bayesian test follows from Equations 3 and 4:

$$\begin{aligned} \mathrm{BF}_{12} &= \frac{p(E_1, E_2|H_1)}{p(E_1, E_2|H_2)} = \\ &= \frac{\prod_i \frac{\frac{n^1_{i*}!}{e^1_{i1}!\dots}\frac{n^2_{i*}!}{e^2_{i1}!\dots}}{\frac{(n^1_{i*}+n^2_{i*})!}{(e^1_{i1}+e^2_{i1})!\dots}} Pólya((e^1_{i1}+e^2_{i1}),\dots|1,\dots)}{\prod_i Pólya(e^1_{i1},\dots|1,\dots)Pólya(e^2_{i1},\dots|1,\dots)} \end{aligned} \quad (9)$$

The Bayes factor for the confusion matrices C_1 and C_2 and the related matrices of errors E_1 and E_2, given in Figure 1, is

$$\mathrm{BF}_{12} = \frac{\mathcal{L}_1}{\mathcal{L}_2} \approx \frac{1.04 \times 10^{-12}}{1.23 \times 10^{-13}} \approx 8.47. \quad (10)$$

According to the guidelines for the interpretation of the value of the Bayes factor reported in Figure 3, the amount of evidence in favour of H_1, i.e. that the pattern of errors of the two confusion matrices is the same, is *substantial*.

A numerically stable implementation of the computation of the Bayes factor, together with some Monte Carlo approximations, are available in Python and Matlab code at https://github.com/emanuele/error_test

III. Experiments

We compare our Bayesian test to two established methods of computing similarity of confusion matrices using the fMRI data from [4]. In this experiment, ten participants passively viewed blocks of images from six natural scene categories (beaches, city streets, forests, highways, mountains, and offices). Each participant was shown eight runs of six blocks of color photographs (CP) and eight runs of six blocks of line drawings (LD) of the same scenes. A linear support vector machine was used to predict the category of scenes in each block in a leave-one-run-out (LORO) cross-validation, separately for CPs and LDs, from fMRI activity in a set of pre-determined regions of interest (ROI). A central question of this study was to determine whether CPs and LDs share the same neural representation of scene categories.

LORO cross-validation resulted in two confusion matrices, one for CPs and one for LDs, for each of the ten participants and for each of seven ROIs: visual areas V1, V2 and V4, the parahippocampal place area (PPA), the retrosplenial cortex (RSC), the lateral occipital complex (LOC), and the fusiform face area (FFA). Here we compute the Bayes factor for the comparison of confusion matrices from CPs and LDs according to Equation 9, separately for each participant. We then compute the joint Bayes factor over all ten participants, which is the product of the individual Bayes factors, because the individual likelihoods can be assumed to be independent:

$$\mathrm{BF}^{\text{joint}}_{12} = \frac{\prod_s \mathcal{L}_{1,s}}{\prod_s \mathcal{L}_{2,s}} = \prod_s \mathrm{BF}_{12,s} \quad (11)$$

where $s = 1, ..., 10$ is the index over individual participants.

ROI	$\mathbf{\log_{10} BF^{joint}_{12}}$	**ECorr**	$\mathbf{NMI \times 10^{-3}}$
V1	-1.47	0.299	8.58
V2	0.355	0.450	12.8
V4	2.87	0.411^{*}	2.60
PPA	1.68	0.631^{**}	31.1
RSC	2.50	0.692^{**}	6.89
LOC	1.37	0.438^{*}	1.04
FFA	-4.05	0.104	0.473

Fig. 4. Joint Bayes factor, error correlation and normalized mutual information for the comparison of confusion matrices for predicting scene category of color photographs and line drawings from fMRI data. $^{*}p < 0.05, ^{**}p < 0.01$.

We compare these values for all seven ROIs with two other methods, both using the mean of the confusion matrices over participants, normalized so that the row-marginals are all 1. Error correlation (ECorr) was computed as the Pearson correlation of the off-diagonal elements of these group-mean confusion matrices for CPs and LDs. Statistical significance of ECorr was established non-parametrically against the null distribution of all error correlations obtained from jointly permuting rows and columns of one of the two confusion matrices.

To compute normalized mutual information (NMI), confusion matrices were interpreted as tables of conditional probablities of predicted labels, given ground truth labels. By computing the appropriate marginal and joint probablities and their respective entropies, we computed the mutual information of the labels predicted for CP, given the labels predicted for LD [2]. Mutual information provides a measure of the reduction of expected uncertainty (entropy) about the category labels of CPs, given that we know the predictions for LDs (and vice versa). Mutual information was normalized to the interval $[0, 1]$.

Figure 4 shows a comparison of results from the three methods for all seven ROIs. There is good agreement between joint BF and ECorr ($r = 0.846, p = 0.0165$; Spearman's rank correlation; see Figure 5), but NMI does not agree well with joint BF ($r = 0.247$) and ECorr ($r = 0.543$). The low agreement of NMI with the other two methods could be due to the inclusion of the diagonal in the computation of NMI. Thereby, decoding accuracy becomes part of NMI and not just the error patterns as for joint BF and ECorr.

The joint BFs for the individual ROIs have plausible scientific interpretations. Area V1, the first processing stage of visual information in cortex, shows strong evidence in favor of

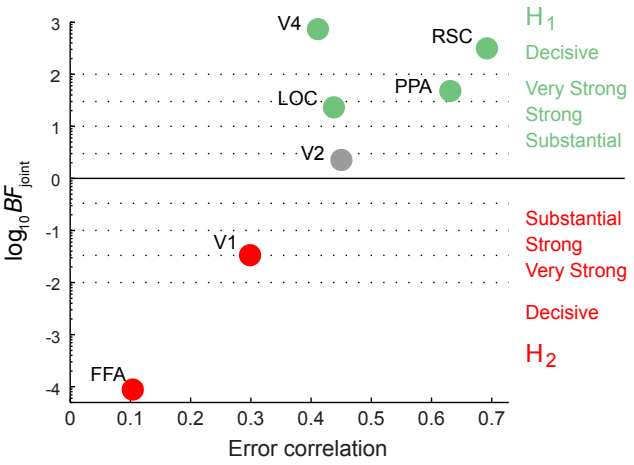

Fig. 5. Comparison of joint Bayes factor and error correlation. Data points are colored green if they support H_1 and red if they support H_2.

H_2 (CPs and LDs result in different error patterns), which is compatible with V1's role as an early filtering stage. As visual information ascends the processing hierarchy, representations of scene category for CP and LD become more similar: inconclusive for V2, very strongly or decisively favoring H_1 (CP and LD result in the same error patterns) for V4, RSC and PPA. Wheras RSC and PPA are visual areas specialized in scene processing, LOC is mostly involved in object recognition and is only indirectly implicated in scene processing. This matches well with the "only" strong evidence in favor of H_1. FFA, on the other hand, is specialized in face processing and not at all in the processing of scenes. Hence we find decisive support in favor of H_2 for the FFA. The joint BF for V4 is somewhat of an outlier: it is the largest value among the seven ROIs, even though error correlation for V4, although significant, is comparably moderate. This disproportionately large joint BF is likely caused by an outlier participant with an unusually high BF for V4.

IV. Discussion

We have introduced a Bayesian test for the dependence or independence of the classification errors from two confusion matrices, including analytical solutions for the computation of the likelihoods of dependence and independence. We have demonstrated the viability of our method using fMRI data relating the neural representation of categories of natural scenes elicited by color photographs to the representation elicited by line drawings of the same scenes. The results of the Bayesian test are in approximate agreement with error correlations. However, unlike error correlation, the Bayesian test offers a clear statistical interpretation in favor of *either* hypothesis. Additional work will be necessary to make this technique a full-fledged tool for the analysis of categorical neuroimaging data, including accounting for multiple comparisons as well as correlations of nearby locations for searchlight analysis.

Appendix
Derivation of $\mathcal{L}_1$

$$
\begin{aligned}
\mathcal{L}_1 &= p(C1, C2|H_1) = \qquad (12)\\
&= \int \prod_i Mul(e^1_{i1}, e^1_{i2}, \ldots | n^1_{i*}, \theta_{i1}, \theta_{i2}, \ldots)\\
&\quad Mul(e^2_{i1}, e^2_{i2}, \ldots | n^2_{i*}, \theta_{i1}, \theta_{i2}, \ldots)\\
&\quad Dir(\theta_{i1}, \theta_{i2}, \ldots | 1, 1, \ldots) d\theta_{i1} d\theta_{i2} \ldots =\\
&= \int \prod_i \frac{n^1_{i*}!}{e^1_{i1}! e^1_{i2}! \ldots} \theta^{e^1_{i1}}_{i1} \theta^{e^1_{i2}}_{i2} \ldots \frac{n^2_{i*}!}{e^2_{i1}! e^2_{i2}! \ldots} \theta^{e^2_{i1}}_{i1} \theta^{e^2_{i2}}_{i2} \ldots\\
&\quad Dir(\theta_{i1}, \theta_{i2}, \ldots | 1, 1, \ldots) d\theta_{i1} d\theta_{i2} \ldots =\\
&= \prod_i \frac{n^1_{i*}!}{e^1_{i1}! e^1_{i2}! \ldots} \frac{n^2_{i*}!}{e^2_{i1}! e^2_{i2}! \ldots} \int \theta^{e^1_{i1}+e^2_{i1}}_{i1} \theta^{e^1_{i2}+e^2_{i2}}_{i2} \ldots\\
&\quad Dir(\theta_{i1}, \theta_{i2}, \ldots | 1, 1, \ldots) d\theta_{i1} d\theta_{i2} \ldots =\\
&= \prod_i \frac{\frac{n^1_{i*}!}{e^1_{i1}! e^1_{i2}! \ldots} \frac{n^2_{i*}!}{e^2_{i1}! e^2_{i2}! \ldots}}{\frac{(n^1_{i*}+n^2_{i*})!}{(e^1_{i1}+e^2_{i1})! (e^1_{i2}+e^2_{i2})! \ldots}}\\
&\quad \int Mul(e^1_{i1} + e^2_{i1}, e^1_{i2} + e^2_{i2} \ldots | n^1_{i*} + n^2_{i*}, \theta_{i1}, \theta_{i2}, \ldots)\\
&\quad Dir(\theta_{i1}, \theta_{i2}, \ldots | 1, 1, \ldots) d\theta_{i1} d\theta_{i2} \ldots =\\
&= \prod_i \frac{\frac{n^1_{i*}!}{e^1_{i1}! e^1_{i2}! \ldots} \frac{n^2_{i*}!}{e^2_{i1}! e^2_{i2}! \ldots}}{\frac{(n^1_{i*}+n^2_{i*})!}{(e^1_{i1}+e^2_{i1})! (e^1_{i2}+e^2_{i2})! \ldots}}\\
&\quad Pólya((e^1_{i1} + e^2_{i1}), (e^1_{i2} + e^2_{i2}), \ldots | 1, 1, \ldots).
\end{aligned}
$$

where

$$
\begin{aligned}
Pólya(\mathbf{z}|\boldsymbol{\alpha}) &= \int Mul(\mathbf{z}|\boldsymbol{\theta}) Dir(\boldsymbol{\theta}|\boldsymbol{\alpha}) d\boldsymbol{\theta} =\\
&= \frac{\Gamma(\sum_i \alpha_i)}{\Gamma(\sum_i z_i + \alpha_i)} \prod_i \frac{\Gamma(z_i + \alpha_i)}{\Gamma(\alpha_i)} \qquad (13)
\end{aligned}
$$

is the Dirichlet compound multinomial (DCM) distribution, also known as multivariate Pólya distribution.

References

[1] D. B. Walther, D. M. Beck, and L. Fei-Fei, "To err is human: correlating fMRI decoding and behavioral errors to probe the neural representation of natural scene categories." in *Visual population codes - Toward a common multivariate framework for cell recording and functional imaging*, N. Kriegeskorte and G. Kreiman, Eds. Cambridge, MA: MIT Press, 2012, pp. 391–415.

[2] D. B. Walther, "Using Confusion Matrices to Estimate Mutual Information between Two Categorical Measurements," in *Pattern Recognition in Neuroimaging (PRNI), 2013 International Workshop on*. IEEE, Jun. 2013, pp. 220–224. [Online]. Available: http://dx.doi.org/10.1109/prni.2013.63

[3] R. E. Kass and A. E. Raftery, "Bayes Factors," *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 773–795, Jun. 1995. [Online]. Available: http://dx.doi.org/10.2307/2291091

[4] D. B. Walther, B. Chai, E. Caddigan, D. M. Beck, and L. Fei-Fei, "Simple line drawings suffice for functional MRI decoding of natural scene categories," *Proceedings of the National Academy of Sciences*, vol. 108, no. 23, pp. 9661–9666, Jun. 2011. [Online]. Available: http://dx.doi.org/10.1073/pnas.1015666108