Optimal Combination of Contrast Estimates

W.D. Penny

Wellcome Trust Centre for Neuroimaging,
University College, London WC1N 3BG, UK.

December 30, 2015

1 Independent Contrasts

Say we have experimental data y_1, and a linear model with design matrix X_1 and observation noise covariance Σ_1 then the regression coefficients are distributed as

$$p(\beta_1) = \mathsf{N}(\beta; m_{\beta_1}, C_{\beta_1}) \tag{1}$$

with mean and covariance given by

$$m_{\beta_1} = \left(X_1^T \Sigma_1^{-1} X_1\right)^{-1} X_1^T \Sigma_1^{-1} y_1 \tag{2}$$

The covariance of these estimates is given by

$$C_{\beta_1} = \left(X_1^T \Sigma_1^{-1} X_1\right)^{-1} \tag{3}$$

The maximum likelihood estimate of β_1 is m_{β_1}.

For data y_2, design X_2 and noise covariance Σ_2 we have equivalent expressions for β_2, m_{β_2} and C_{β_2}.

The contrast vectors c_1 and c_2 are then used to capture the experimental effects of interest, w_1 and w_2, from each fitted model. The mean and covariance of the estimated effects are then given by

$$\begin{aligned} m_{w_1} &= c_1^T m_{\beta_1} && (4) \\ C_{w_1} &= c_1^T C_{\beta_1} c_1 && (5) \end{aligned}$$

If we have scalar effects then m and C here are also scalars. I'll write $m_1 \equiv m_{w_1}$ and $\lambda_1 \equiv 1/C_{w_1}$. Similar expressions exist for contrast 2 and model 2.

The optimal way to combine the two contrasts estimates is then

$$\begin{aligned} \lambda &= \lambda_1 + \lambda_2 && (6) \\ m &= \frac{\lambda_1}{\lambda} m_1 + \frac{\lambda_2}{\lambda} m_2 \end{aligned}$$

2 Dependent Contrasts

Say we have experimental data y, and a linear model with design matrix X and observation noise covariance Σ. Note that y may contain two time series of interest y_1 and y_2. The matrix Σ may describe (error) correlations among these time series.

The regression coefficients are distributed as

$$p(\beta) = \mathsf{N}(\beta; m_\beta, C_\beta) \tag{7}$$

with mean and covariance given by

$$m_\beta = \left(X^T \Sigma^{-1} X\right)^{-1} X^T \Sigma^{-1} y \tag{8}$$

$$C_\beta = \left(X^T \Sigma^{-1} X\right)^{-1} \tag{9}$$

The contrast matrix C is then used to capture the experimental effects of interest, w. The mean and covariance of the estimated effects are then given by

$$m_w = C^T m_\beta \tag{10}$$

$$C_w = C^T C_\beta C \tag{11}$$

If C has two rows, with the first row corresponding to c_1 and the second row to c_2 then our effects of interest $w = [w_1, w_2]^T$. The mean m_w is then a 2-by-1 vector, and C_w is a 2-by-2, with the off-diagonal element capturing covariance (dependence) in our contrast estimates. Generally, C may have N rows.

I'll write $m_w \equiv [m_1, m_2, ..m_N]^T$ and $\Lambda \equiv C_w^{-1}$. The optimal way to combine the estimates in m_w is

$$m = \left(1_N^T \Lambda 1_N\right)^{-1} 1_N^T \Lambda m_w \tag{12}$$

where 1_N is a column vector of N ones. This result stems from treating m_w as containing entries that vary around a common mean m with precision Λ. The fact we are treating m as a common mean is specified by setting the (higher-level) 'design matrix' to 1_N.

To see that this result makes sense, consider the case with $N = 2$ and λ_1, λ_2 and λ_{12} are diagonal and off-diagonal entries in Λ. Then we have

$$m = \frac{\lambda_1 m_1 + \lambda_2 m_2 + \lambda_{12} m_1 + \lambda_{12} m_2}{\lambda_1 + \lambda_2 + \lambda_{12}} \tag{13}$$

Reassuringly, if the contrasts are independent we have $\lambda_{12} = 0$ and we have the same result as in equation 6.