

Bayesian odds ratio of two multinomials

Luis Pedro Coelho

December 17, 2012

The equations in this derivation have not been checked by automatic software. Unfortunately, I cannot think of a way of checking them easily as I have not found a piece of software which understood Dirichlet integrals in k dimensions.

We are given two datasets \mathcal{D}_1 and \mathcal{D}_2 , which are completely summarised as the output of a multinomial of size k . The question is: *were these two datasets produced by a single multinomial or two different multinomial?* We will call the single multinomial hypothesis \mathcal{H}_1 and the two multinomial hypothesis \mathcal{H}_2 .

If we call \mathcal{D} the complete data, we are enquiring after:

$$\frac{P(\mathcal{H}_1|\mathcal{D})}{P(\mathcal{H}_2|\mathcal{D})}. \quad (1)$$

If we assume a prior with the same value for both hypothesis, this is equivalent to

$$\frac{P(\mathcal{D}|\mathcal{H}_1)}{P(\mathcal{D}|\mathcal{H}_2)}. \quad (2)$$

We will also assume that the two datasets are independent in the case of a single multinomial, i.e.,

$$\frac{P(\mathcal{D}|\mathcal{H}_1)}{P(\mathcal{D}|\mathcal{H}_2)} = \frac{P(\mathcal{D}_1, \mathcal{D}_2|\mathcal{H}_1)}{P(\mathcal{D}_1, \mathcal{D}_2|\mathcal{H}_2)} = \frac{P(\mathcal{D}_1|M)P(\mathcal{D}_2|M)}{P(\mathcal{D}_1, \mathcal{D}_2|M)}, \quad (3)$$

where M is the multinomial model. We now consider what form $P(\mathcal{D}_i|M)$ takes. The data is simply a vector of counts, which we will call \vec{x} . We need to integrate over all possible multinomials, which are in turn defined by a vector $\vec{\theta}$ in the simplex defined by $\sum_i \theta_i = 1$. We use a Dirichlet prior, which needs a new parameter $\vec{\alpha}$.

$$P(\vec{c}|\alpha, M) = \int P(\vec{c}|\vec{\theta}, M) P(\vec{\theta}|M) d\theta \quad (4)$$

$$= \frac{1}{D(\vec{\alpha})} \int (\Pi_i \theta_i^{c_i}) (\Pi_i \theta_i^{\alpha_i - 1}) d\theta \quad (5)$$

$$= \frac{1}{D(\vec{\alpha})} \int \Pi_i \theta_i^{c_i + \alpha_i - 1} d\theta \quad (6)$$

$$= \frac{D(\vec{c} + \vec{\alpha})}{D(\vec{\alpha})}, \quad (7)$$

where D is the Dirichlet normalizing constant:

$$D(\vec{\alpha}) = \frac{\prod_i \Gamma(\alpha_i)}{\Gamma(\sum_i \alpha_i)}, \quad (8)$$

which is also sometimes called the multinomial Beta (but too many functions are already called Beta).

Back to (3). For convenience, I will call the vector of counts from \mathcal{D}_1 , \vec{x} and that from \mathcal{D}_2 , \vec{y} . (3) expands to

$$\frac{D(\vec{x} + \alpha)}{D(\vec{\alpha})} \cdot \frac{D(\vec{y} + \alpha)}{D(\vec{\alpha})} \cdot \frac{D(\vec{\alpha})}{D(\vec{x} + \vec{y} + \alpha)} = \frac{D(\vec{x} + \vec{\alpha})D(\vec{y} + \vec{\alpha})}{D(\vec{x} + \vec{y} + \vec{\alpha})D(\vec{\alpha})}. \quad (9)$$

With another another assumption, namely that $\alpha_i = 0$ (and $D(\vec{\alpha}) = 1$, which implies an improper prior), we can get a nicer expression:

$$\frac{P(\mathcal{D}|\mathcal{H}_1)}{P(\mathcal{D}|\mathcal{H}_2)} = \frac{D(\vec{x})D(\vec{y})}{D(\vec{x} + \vec{y})}. \quad (10)$$

If we further define the Beta function B as:

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x + y)}, \quad (11)$$

then (10) become:

$$\frac{P(\mathcal{D}|\mathcal{H}_1)}{P(\mathcal{D}|\mathcal{H}_2)} = \frac{D(\vec{x})D(\vec{y})}{D(\vec{x} + \vec{y})} = \prod_i B(x_i, y_i). \quad (12)$$