

# PRIORID: Statistical Calculations

## 1 Statistical details

Once the set of non-genetic variables to be evaluated is established, each combination of values of the variables defines a cell. First, to simplify the notation, we number the cells from 1 to  $m$ , being  $m$  the total number of cells. Let us call  $N_j$  the total number of possible victims belonging to the  $j$  cell. We consider a Dirichlet distribution as the prior probabilities for obtaining every cells.

$$f(\theta_1, \theta_2, \dots, \theta_m) = \frac{\Gamma(\alpha_1 + \alpha_2 + \dots + \alpha_j + \dots)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\dots\Gamma(\alpha_j)\dots} \theta_1^{\alpha_1} \theta_2^{\alpha_2} \dots \theta_m^{\alpha_m}, \quad (1)$$

The hyperparameters are calibrated in order to be consistent with the prior knowledge, thus

$E(\theta_j) = \frac{N_j}{N}$  for all  $j = 1, 2, \dots, m$  (where  $E(\theta_j)$  means expected value or "average" value of  $\theta_j$ ). Therefore, in the instance of the knowledge prior of the data of the already-solved cases, all the individuals have the same chance of corresponding with the skeletal remains  $S$ . As for the Dirichlet distribution, the expected value of the variable  $\theta_j$  is  $\frac{\alpha_j}{\alpha_0}$  being  $\alpha_0 = \alpha_1 + \alpha_2 + \dots + \alpha_m$ , then what must be satisfied are  $m$  conditions, one for each cell. Assuming an extra condition for the  $k$  cell such as  $N_k$  is maximum, such as  $\frac{\sqrt{Var(\theta_k)}}{E(\theta_k)} = 1$ , the solution for the  $\alpha_0$  hyperparameter is  $\alpha_0 = \frac{N}{N_k} - 2$ , and the expression for each  $\alpha_j$  hyperparameter corresponding to the  $j$  cell is:

$$\alpha_j = \frac{N_j}{N} \left( \frac{N}{N_k} - 2 \right) \quad \forall j \quad (2)$$

Finally, using that Dirichlet and Multinomial distributions are conjugate distributions, then if the likelihood is a multinomial and the prior probability of the parameters is a Dirichlet, then the Posterior distribution of the parameters  $(\theta_1, \theta_2, \dots, \theta_m | Data)$  is also a Dirichlet distribution but with other parameters, which are going to be a function of the parameters of the prior and the *Data*, such as  $\alpha'_j = \alpha_j + I_j$ , in general for cell  $j$ , being  $I_j$  the total number of already-solved cases of the massive event associated to  $j$  cell. This means that in the ideal problem the probability that the outcome result of  $j$  cell will be

$$\frac{\alpha_j + I_j}{\alpha_0 + n_s} = \frac{\frac{N_j}{N} \left( \frac{N}{N_k} - 2 \right) + I_j}{\frac{N}{N_k} - 2 + n_s} \equiv \hat{\theta}_j^{\text{post}}$$

being  $n_s$  the total number of already-solved cases.

In order to construct a probabilistic ranking for each possible victim we assume that within a particular cell, all of the people which are unidentified have the same chance of corresponding with the skeletal remain of the same event. Thus the probability of a particular individual whose non-genetic variables values are associated to cell  $j$  is

$$P = \frac{\hat{\theta}_j^{\text{post}}}{N_j}$$

Odds probability are calculated as  $ODDS = P/(1 - P)$ .

### *References*

A. O'Hagan, C.E. Buck, A. Daneshkhah, J.R. Eiser, P.H. Garthwaite, D.J. Jenkinson, J.E. Oakley, T. Rakow, Uncertain Judgements: Eliciting Experts' Probabilities (Statistics in Practice). Wiley, Nueva York, EE.UU. (2006).