Cálculos estadísticos

Una vez definidas las variables no genéticas a ser evaluadas, cada combinación de valores de estas variables define un casillero. En primer lugar, para simplificar la notación, numeramos los casilleros desde el 1 hasta el m, siendo m el número total de casilleros posibles. Llamemos N_j al número total de víctimas posibles pertenecientes al casillero j.

Consideramos una distribución del tipo Dirichlet como la probabilidad a priori (de los casos resueltos) para cada casillero de corresponderse con el resto óseo del evento considerado:

$$Priori\left((\theta_1, \theta_2, ..., \theta_j, ...)|(\alpha_1, \alpha_2, ..., \alpha_j, ...)\right) = \frac{\Gamma(\alpha_1 + \alpha_2 + ... + \alpha_j + ...)}{\Gamma(\alpha_1)\Gamma(\alpha_2)...\Gamma(\alpha_j)...}\theta_1^{\alpha_1}\theta_2^{\alpha_2}...\theta_j^{\alpha_j}... \tag{1}$$

Los hiperparámetros son calibrados para que sean consistentes con el conocimiento a priori de los casos ya resueltos del mismo evento, es decir:

$E(\theta j) = \frac{N_j}{N}$ para todo casillero $j = 1, 2..., m$ (donde$E(\theta_j)$ significa esperanza o valor esperado of θ_j). Por lo tanto, en la instancia de conocimiento (a priori de los casos resueltos) todos los individuos tienen la misma chance de corresponderse con los restos óseos S. Como para la distribución Dirichlet el valor esperado de la variable θ_j es α_j/α_0 siendo $\alpha_0 = \alpha_1 + \alpha_2 + ... + \alpha_m$, entonces se deben cumplir m condiciones, una para cada casillero. Asumiendo una condición extra para el casillero k para el que N_k toma el máximo valor (es decir, es el casillero más poblado) por la cual pedimos que $\frac{\sqrt{Var(\theta_k)}}{E(\theta_k} = 1$, entonces la solución para el hiperparámetro α_0 es $\alpha_0 = \frac{N}{N_k} - 2$, y la expresión para cada hiperparámetro α_i correspondiente al casillero i es:

$$\alpha_i = \frac{N_i}{N}\left(\frac{N}{N_k} - 2\right) \quad \forall\ i$$

Finalmente, usando que la distribucuión de Dirichlet y la Multinomial son distribuciones conjugadas, entonces si la verosimilitud es una multinomial y la priori una Dirichlet, como asumimos en este problema, entonces la Posterior $Post\left((\theta_1, \theta_2...\theta_m)|Datos, \mathcal{H}\right)$ será también una distribución Dirichlet pero con otros hiperparámetros, que son una función de los hiperparámetros originales y de los datos ($Datos$) de los casos resueltos del evento:

$$\alpha_j^{'} = \alpha_j + n_j$$

siendo n_j el número total de casos resueltos del evento masivo asociados al casillero j. Esto significa que en el problema ideal la probabilidad de que el nuevo resultado coresponda con el casillero j es:

$$\frac{\alpha_j + n_j}{\alpha_0 + n} = \frac{\frac{N_j}{N}\left(\frac{N}{N_k} - 2\right) + n_j}{\frac{N}{N_k} - 2 + n} \equiv \theta_j^{post}$$

siendo n el número total de casos ya resueltos del evento considerado.

Para construir un ranking probabilístico para cada posible víctima asumimos que dentro de un dado casillero, todos los individuos aún no identificados tienen la misma chance de corresponderse con el resto óseo o identificado del evento considerado. De esta forma, la probabilidad para un individuo particular cuyas variables no genéticas están asociadas al casillero j es

$$P = \frac{\theta_j^{post}}{N_j - n_j}$$

Los cocientes de Odds se calculan como $ODDS = P/(1 - P)$.

Referencias

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