

Disease determination algorithm

24 January 2016

Patrick D. Emond

The probability of disease status of each individual is determined by the following equation:

$$P(i) = \frac{1}{1 + e^{-\eta(i)}}$$

where i is a particular individual and η is the sum of all weighted factors:

$$\eta(i) = \sum_f W_f Z_f(i)$$

where f represents each disease factor, W_f is the weight for the disease factor f and $Z_f(i)$ is the normalized value for the factor f , for the individual i :

$$Z_f(i) = \frac{X_f(i) - \mu_f}{\sigma_f}$$

where $X_f(i)$ is the individual's value for factor f , and μ_f and σ_f is the mean and standard deviation of all values for factor f . In one equation this resolves to:

$$P(i) = \left\{ 1 + \exp \left[- \sum_f W_f \frac{X_f(i) - \mu_f}{\sigma_f} \right] \right\}^{-1}$$

The six factors include population, income, disease risk, age, sex and pocketing:

$$\begin{aligned} X_{population} &\equiv N_{household} \\ X_{income} &\equiv -\ln \mathcal{N} \left\{ F_\mu()|_{\square}, F_\sigma()|_{\square} \right\} \\ X_{risk} &\equiv \mathcal{N} \left\{ F_\mu()|_{\square}, F_\sigma()|_{\square} \right\} \\ X_{age} &\equiv \begin{cases} 0, & \text{if child} \\ 1, & \text{if adult} \end{cases} \\ X_{sex} &\equiv \begin{cases} 0, & \text{if female} \\ 1, & \text{if male} \end{cases} \\ X_{pocket} &\equiv \sum_{pocket} \begin{cases} e^{-d_p}, & \text{if exponential} \\ d_p^{-2}, & \text{if inverse square} \\ e^{-d_p^2}, & \text{if Gaussian} \end{cases} \end{aligned}$$

where $\ln \mathcal{N}(\mu, \sigma)$ and $\mathcal{N}(\mu, \sigma)$ are the log-normal and normal distributions for mean μ and standard deviation σ , $F()$ is a spacial functions of the form $F() = b_{00} + b_{10}x + b_{01}y + b_{20}x^2 + b_{02}y^2 + b_{11}xy$, $F()|_{\square}$ is the function $F()$ evaluated at the centroid of a tile and d_f is the distance to the pocket p divided by a programmable scaling constant.

By default this algorithm will result in the mean disease prevalence always being near 0.5. However, it is possible to change the resulting mean to a target mean, τ by shifting the function for P such that it crosses the Y-axis at τ . First we must solve P for η :

$$\eta = -\ln\left(\frac{1}{P} - 1\right)$$

then add the value of η for $P = \tau$ from the function for $\eta(i)$:

$$\eta(i) = \sum_f W_f Z_f(i) - \ln\left(\frac{1}{\tau} - 1\right)$$

Finally, the above adjustment to η will result in a mean prevalence that follows an arc-sin curve. In order to get the desired mean prevalence the value for τ must be adjusted by an ad-hoc function:

$$\tau_o = \frac{1}{2} \left\{ \sin\left[\frac{\pi}{2}(2\tau - 1)\right] + 1 \right\}$$