S3 Text. Improving estimates by sampling to find presymptomatic infection

As described in the Methods, testing asymptomatic individuals to find presymptomatic infection can make the probability distribution for the current size of the outbreak more precise. The probability of correctly identifying presymptomatic infection when an infected individual is tested is denoted by p_d. If E_n is the number of presymptomatic infected individuals correctly identified as infected out of a sample of n asymptomatic individuals, and there are m asymptomatic individuals (including susceptibles) in total, then

$$P(E = e \mid E_n = k) = \frac{P(E_n = k \mid E = e)P(E = e)}{\sum_{j=k}^{m} P(E_n = k \mid E = j)P(E = j)},$$

valid for $e = k, k+1, \ldots, m$, where the final term of the numerator, and the final term of the denominator, can be calculated described in the previous section (i.e. according to equation (S1) in Text S1, or in the case where the incubation period is longer than the latent period, from the posterior from RJMCMC). The first terms of the numerator and denominator are given by

$$P(E_n = k \mid E = j) = \begin{cases} \dfrac{\binom{j}{k}\binom{m-j}{n-k}}{\binom{m}{n}} & \text{for } p_d = 1, \\[2em] \displaystyle\sum_{l=\max(0,n-k-m+j)}^{\min(j,n)-k} \dfrac{\binom{j}{k+l}\binom{m-j}{n-k-l}\binom{k+l}{k}p_d{}^k(1-p_d)^l}{\binom{m}{n}} & \text{for } p_d \in [0,1), \end{cases}$$

where the second expression is valid whenever $\min(j, n) - k \geq \max(0, n-k-m+j)$, and is zero otherwise.