Supplementary material to "Estimating the Attack Ratio of Dengue Epidemics under Time-varying Force of Infection using Aggregated Notification Data"

Flavio Codeço Coelho and Luiz Max de Carvalho

February 19, 2015

A remark on prior distributions and tail behaviour of the distribution of R_t

There are a number of approaches to deriving the distribution of R_t . Alternatively to the approach described in the main text [1], one could use the conditional distribution of R_t on Y_{t+1} and Y_t as defined in equation A7 of Nishiura et al. [2]:

$$f_R(R_t) = (Y_t R_t)^{Y_{t+1}} e^{-Y_t R_t} \tag{1}$$

Noticing the kernel of (1) is that of a gamma distribution with $a_2 = Y_{t+1} + 1$ and $b_2 = Y_t$, we obtain a proper density from which to construct $c_\alpha(R_t)$, simply by computing the appropriate quantiles of said distribution. This density is

$$f_N(R_t|a_2, b_2) = \frac{b_2^{a_2}}{\Gamma(a_2)} R_t^{a_2-1} e^{-b_2 R_t} \tag{2}$$

In order to decide which approach to take, it may be of use analysing the tail behaviour of the derived distributions for R_t . Consider the case of using a flat $Uniform(0, 1)$ prior for θ_t. With $a_0 = b_0 = 1$, $a_1 = a_2$ and $b_1 = b_2 + 1$. The beta prime (inverse beta distribution) will have heavier tails compared to the conditional distribution proposed by [2], thus providing more conservative confidence/credibility intervals. To see that one needs simply take the ratio of the Beta prime and Gamma (unnormalized) densities and evaluate the limit as R_t goes to infinity:

$$\lim_{R_t \to \infty} \frac{f_P(R_t|a_1, b_1)}{f_N(R_t|a_2, b_2)} = \lim_{R_t \to \infty} \frac{e^{Y_t R_t}}{(1 + R_t)^{Y_t + Y_{t+1} + 2}} = \infty \tag{3}$$

Finally, note that we deliberately construct $c_\alpha(R_t)$ as a equal-tailed $100\alpha\%$ credible set, rather than a less conservative highest posterior density (HPD) interval.

As a side note, the Bayesian approach presented in this paper will give similar results to orthodox confidence intervals [3] and [4] for Y_{t+1} and $Y_t >> 1$. Under the flat uniform prior for θ_t, the Bayesian posterior credibility interval is nearly

indistinguishable from the confidence interval proposed by Clopper & Pearson (1931) [4] for $Y_{t+1}, Y_t > 20$. Note also that the uniform prior ($Beta(1,1)$) for θ_t constitutes a poor prior choice mainly because the induced distribution for R_t is only well-defined for $b_0 > 2$.

An advantage of the Bayesian approach is that one can devise prior distributions for θ_t taking advantage of the intuitive parametrization and flexibility of the beta family of distributions. Prior elicitation can also be done for R_t and the hyper-parameters directly plugged into the prior for θ_t. One can, for example, choose prior mean and variance for R_t and find a_0 and b_0 that satisfy those conditions. Let m_0 and v_0 be the prior expectation and variance for R_t. After some tedious algebra one finds

$$a_0 = \frac{m_0 v_0 + m_0^3 + m_0^2}{v_0} \tag{4}$$

$$b_0 = \frac{2v_0 + m_0^2 + m_0}{v_0} \tag{5}$$

If one wants only to specify m_0 and the coefficient of variation $c = \sqrt{v_0}/m_0$ for R_t *a priori*, some less boring algebra gives:

$$a_0 = \frac{m_0^3 c^2 + m_0^3 + m_0^2}{m_0^2 c^2} \tag{6}$$

$$b_0 = \frac{2m_0^2 c^2 + m^2 + m}{m_0^2 c^2} \tag{7}$$

This approach thus makes it possible to incorporate epidemiological knowledge about disease Biology (e.g. the magnitude of R_0) into the computation of R_t . This may prove particularly important when disease counts are low and/or close to the detection threshold. We provide an R script to perform the above elicitation at `https://github.com/fccoelho/paperLM1/blob/master/R/elicit_Rt_prior.R`.

References

[1] Ederer, F. & Mantel, N. Confidence limits on the ratio of two poisson variables. *American Journal of Epidemiology* **100**, 165–167 (1974).

[2] Nishiura, H., Chowell, G., Heesterbeek, H. & Wallinga, J. The ideal reporting interval for an epidemic to objectively interpret the epidemiological time course. *J R Soc Interface* **7**, 297–307 (2010).

[3] Wilson, E. B. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association* **22**, 209–212 (1927).

[4] Clopper, C. & Pearson, E. S. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 404–413 (1934).