Technical Report
Quantifying uncertainty about the estimates of the effective reproductive number R_t

Luiz Max F. de Carvalho

Program for Scientific Computing (PROCC)

Oswaldo Cruz Foundation

September 1, 2014

Abstract

We derive confidence/credibility bounds for the uncertainty about the estimate of the effective reproductive number R_t under the assumption that reported disease counts are Poisson distributed. It is shown that one can construct a conservative confidence interval for R_t using the induced probability distribution that results from the approach of Ederer and Mantel (1974). Also, we derive the confidence bounds using the conditional distribution of R_t on the observed disease counts (Nishiura et al., 2010), which presents lighter tails.

1 Background

In monitoring of infectious diseases, it is important to assess whether the incidence of a a particular disease is increasing significantly, in order to decide to take preventive measures. The effective reproductive number at time t, R_t , can be understood as a real-time estimate of the basic reproductive number (R_0) and is defined as the average number of secondary cases per primary case at time t. Let Y_t be the number of reported disease cases for a particular time $t \in (0, T)$. Nishiura et al. (2010) extend the theory developed by Stallybrass et al. (1931) and propose to estimate R_t as

$$R_t = \left(\frac{Y_{t+1}}{Y_t}\right)^{1/n} \tag{1}$$

where n is taken to be the ratio between the length of reporting interval and the mean generation time of the disease. Here we are interested in the simpler case $n = 1$. If R_t is to be used as a decision tool, one needs to be able to quantify the uncertainty about estimate in equation 1. Here we detail how to obtain confidence intervals for R_t , exploring two approaches under the assumption that the counts Y_t are Poisson distributed for all t.

2 Approaches

2.1 Ederer & Mantel (1974)

First we explore the approach of Ederer and Mantel (1974), whose objective is to obtain confidence intervals for the ratio of two Poisson counts. Let $Y_t \sim Poisson(\lambda_t)$ and $Y_{t+1} \sim Poisson(\lambda_{t+1})$ and define $S = Y_t + Y_{t+1}$. The authors then proceed to note that by conditioning on the sum S

$$Y_{t+1}|S \sim Binomial(S, \theta_t) \tag{2}$$

$$\theta_t = \frac{\lambda_{t+1}}{\lambda_t + \lambda_{t+1}} \tag{3}$$

Let $c_\alpha(\theta_t) = \{\theta_t^{(L)}, \theta_t^{(U)}\}$ be such that $Pr(\theta_t^{(L)} < \theta_t < \theta_t^{(U)}) = \alpha$. Analogously, define $c_\alpha(R_t) = \{R_t^{(L)}, R_t^{(U)}\}$ such that $Pr(R_t^{(L)} < R_t < R_t^{(U)}) = \alpha$. Ederer and Mantel (1974) show that one can construct a $100\alpha\%$ confidence interval for R_t by noting that

$$R_t^{(L)} = \frac{\theta_t^{(L)}}{(1 - \theta_t^{(L)})} \tag{4}$$

$$R_t^{(U)} = \frac{\theta_t^{(U)}}{(1 - \theta_t^{(U)})} \tag{5}$$

Many authors have derived $c_\alpha(\theta_t)$ following orthodox approaches such as those of Wilson (1927) and Clopper and Pearson (1934), mainly for simplicity reasons. We choose instead to take a Bayesian approach and use the $100\alpha\%$ posterior credibility interval for θ_t as $c_\alpha(\theta_t)$. If we choose the conjugate beta prior with parameters a_0 and b_0 for the binomial likelihood in (2), the posterior distribution for θ_t is

$$p(\theta_t|Y_{t+1}, S) \sim Beta(Y_{t+1} + a_0, Y_t + b_0) \tag{6}$$

Equation (6) tells us that the induced distribution of R_t is a beta prime with parameters $a_1 = Y_{t+1} + a_0$ and $b_1 = Y_t + b_0$. The density of this induced distribution $f_P(R_t)$ is then

$$f_P(R_t|a_1, b_1) = \frac{\Gamma(a_1 + b_1)}{\Gamma(a_1)\Gamma(b_1)} R_t^{a_1 - 1}(1 + R_t)^{-(a_1 + b_1)} \tag{7}$$

Thus, the expectation of R_t under f_P is $a_1/(b_1 - 1)$ and its variance is $a_1(a_1 + b_1 - 1)/((b_1 - 2)(b_1 - 1)^2)$.

Since when $R_t > 1$ indicates sustained transmission, one may be interested in computing the probability of this event. By noting that

$$Pr(R_t > 1) = 1 - \int_0^1 f_P(r)dr \tag{8}$$

$$= 1 - Pr(\theta_t < \frac{1}{2}) \tag{9}$$

one can avoid dealing with the density in (7) directly.

2.2 Nishiura et al. (2010)

An alternative approach is to use the conditional distribution of R_t on Y_{t+1} and Y_t as defined in equation A7 of Nishiura et al. (2010):

$$f_R(R_t) = (Y_t R_t)^{Y_{t+1}} e^{-Y_t R_t} \tag{10}$$

Noticing the kernel of (10) 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

$$f_N(R_t|a_2, b_2) = \frac{b_2^{a_2}}{\Gamma(a_2)} R_t^{a_2-1} e^{-b_2 R_t} \tag{11}$$

Once again $Pr(R_t > 1)$ can be easily computed using

$$P(R_t > 1) = 1 - F_N(1) = 1 - \frac{1}{\Gamma(a_2)} \gamma(a_2, b_2) \tag{12}$$

where $F_N(R_t)$ is the distribution function of R_t and $\gamma(\cdot)$ is the incomplete gamma function.

2.3 A remark on prior distributions and tail behaviour

In order to decide which approach to take, it may be of use analysing the variance and tail behaviour of the derived distributions for R_t . Consider the case of using a flat $Uniform(0,1)$ prior for θ_t in Section 2.1. 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 in (11), thus providing more conservative confidence/credibility intervals. As a side note, the Bayesian approach presented in Section 2.1 will give similar results to those of Wilson (1927) and Wilson (1927) 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 and Pearson (1934) for $Y_{t+1}, Y_t > 20$. Note that the $Beta(1,1)$ uniform prior for θ_t constitutes a poor prior choice mainly because the induced distribution for R_t is only well-defined for $b_0 > 2$.

The main advange of the Bayesian approach is that one can devise prior distributions for θ_t taking advantage of the intuitive parametrisation and flexibility of the beta family of distributions. Prior elicitation can also be done for R_t and the hyperparameters directly plugged into the prior for θ_t. One can, for example, choose a priori 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

$$a_0 = \frac{m_0 v_0 + m_0^3 + m_0^2}{v_0} \tag{13}$$

$$b_0 = \frac{2v_0 + m_0^2 + m_0}{v_0} \tag{14}$$

If one wants only to specify m_0 and a coefficient of variation c [1] for R_t *a priori*, some less boring algebra gives:

$$a_0 = \frac{m_0^3 c^2 + m_0^3 + m_0^2}{m_0^2 c^2} \tag{15}$$

$$b_0 = \frac{2m_0^2 c^2 + m^2 + m}{m_0^2 c^2} \tag{16}$$

---

[1]$c = \sqrt{v_0}/m_0$.

3 Preliminary results – Dengue notifications in Rio de Janeiro

We now show the application of the methods described above to dengue notification data in the city of Rio de Janeiro. The data consists of weekly dengue notifications for 10 Health Planning Areas (APS, in Portuguese) for the period 2010-2014. Figure 1 shows the data. We compute $c_\alpha(R_t)$ for the total (sum) number of notifications using the two approaches described, using a generation time of 3 weeks and choose a_0 and b_0 such that $E[R_t] = 1/2$ and $Var(R_t) = 2$. The resulting estimates and 95% confidence bounds are presented in Figure 2. Finally, we compute $Pr(R_t > 1)$ using the two distributions discussed here (Figure 3).

4 Final Remarks

Both strategies presented here seem appropriate to the quantification of uncertainty about R_t. The distribution obtained by combining the approach of Ederer and Mantel (1974) with Bayesian inference for θ_t, however, possesses heavier tails and thus offers more conservative bounds. Additionally, the ability to specify prior information on R_t based on expert knowledge may prove useful.

The code to implement the methods discussed and also to generate this report here can be found at https://github.com/maxbiostat/CODE/tree/master/Alerta.

5 Acknowlegements

I am grateful to Dr. Leonardo Bastos for useful discussions and for spotting a bug in my earlier code.
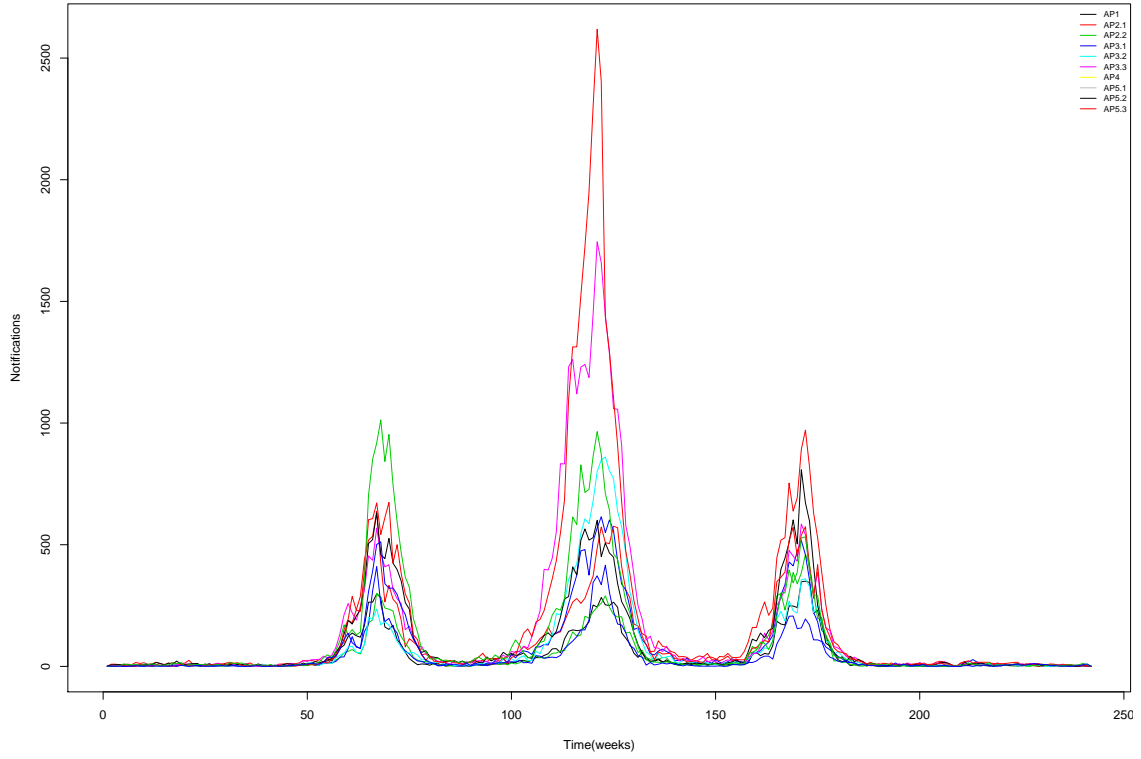
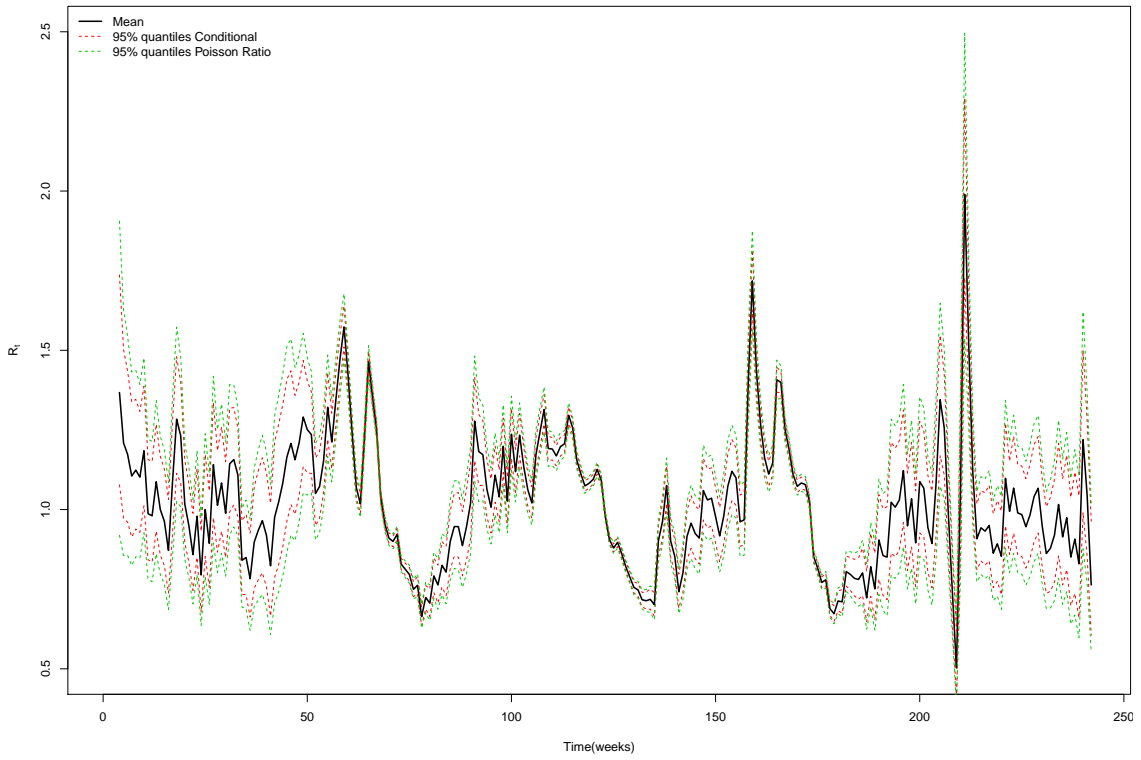Figure 1: Total number of dengue notifications per HPA (APS)

Figure 2: R_t estimates for the total notifications and confidence intervals obtained using the two proposed methods.
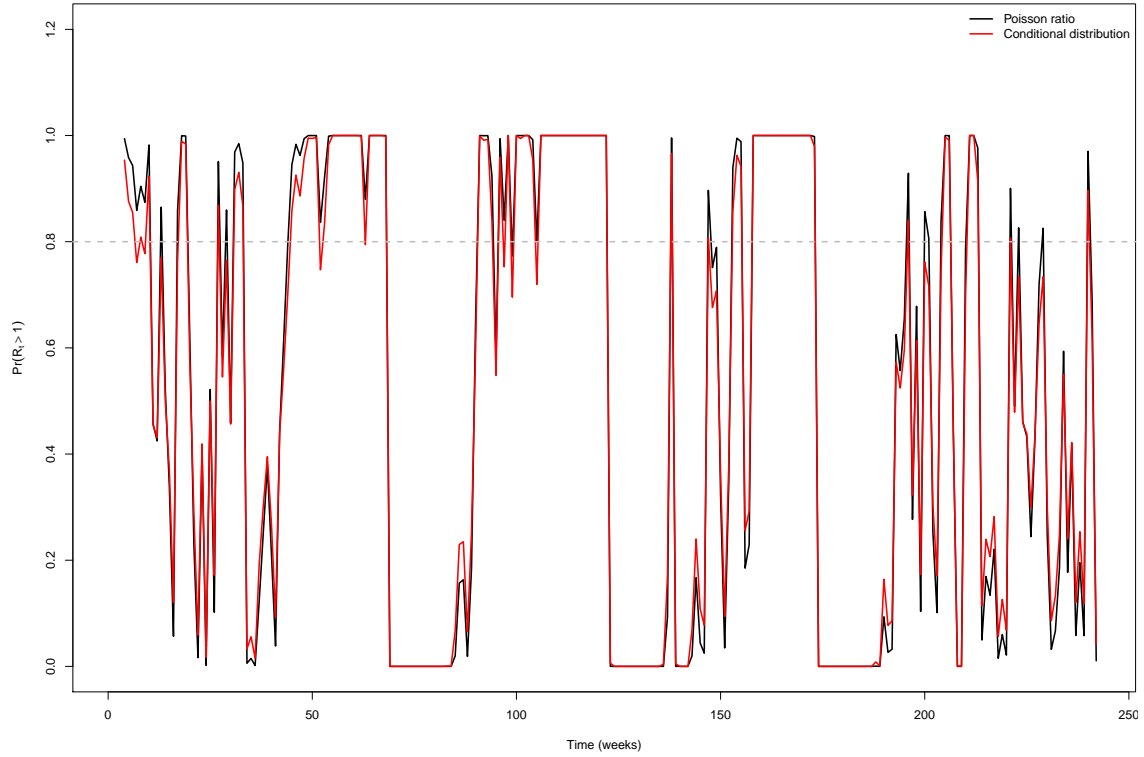
Figure 3: Probability that $R_t \geq 1$. The interrupted gray line marks the arbitrary value of 0.80.

References

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

Ederer, F. and Mantel, N. (1974). Confidence limits on the ratio of two poisson variables. *American Journal of Epidemiology*, 100(3):165–167.

Nishiura, H., Chowell, G., Heesterbeek, H., and Wallinga, J. (2010). The ideal reporting interval for an epidemic to objectively interpret the epidemiological time course. *J R Soc Interface*, 7(43):297–307.

Stallybrass, C. O. et al. (1931). The principles of epidemiology and the process of infection. *The Principles of Epidemiology and the Process of Infection.*

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158):209–212.