

Methodenbeschreibung für die Schätzung von epidemiologischen Parametern des COVID19 Ausbruchs, Österreich

Lukas Richter(1,2)

Daniela Schmid(1)

Ernst Stadlober(2)

1 Abteilung Infektionsepidemiologie & Surveillance, AGES

2 Institut für Statistik, Technische Universität Graz

Einleitung

Für die deskriptive Analyse eines Ausbruchs bzw. einer Epidemie einer Infektionskrankheit benützt man epidemiologische Kennzahlen wie die Anzahl der neu aufgetretenen Fälle bzw. die Inzidenz bezogen auf die Bevölkerung, die zeitliche Verteilung (Epidemiekurve) nach Erkrankungstag der Fälle die regionale Verteilung nach Wohnort der Fälle, sowie die demographischen Merkmale der Fälle, Alters- und Geschlechtsverteilung. Andererseits benützt man auch Kennzahlen, die mit Hilfe statistischer Methoden geschätzt werden können. Wir stellen im Folgenden drei dieser Parameter vor, die Fall-Steigerungsrate, Fall-Verdopplungszeit und die effektive Reproduktionszahl, und beschreiben wie diese berechnet werden können.

Methodenbeschreibung

Steigerungsrate und Verdopplungszeit

Am Beginn eines Ausbruchs geht man davon aus, dass die zeitliche Entwicklung der Fallzahl von inzidenten Fällen einer Poisson Verteilung folgt. Seien also y_t die Anzahl der inzidenten Fälle am Tag t , dann schreiben wir $y_t \sim Poi(\lambda_t)$ und es gilt für den Erwartungswert von y_t , $E(y_t) = \lambda_t$. Der Parameter λ_t wird in der Epidemiologie auch *force of infection* genannt. Wir können λ_t mit dem folgenden linearen Modell beschreiben:

$$\ln(\lambda_t) = r t + b + \varepsilon \tag{1}$$

wobei ε der unbekannte Zufallsfehler ist.

Dabei ist der Regressionskoeffizient, r , die **durchschnittliche tägliche Steigerungsrate** und b ist die Konstante. Die durchschnittliche tägliche Steigerungsrate in Prozent ist dann $(r - 1) \times 100$. Das *R*-Paket **incidence** ist eines der Pakete die die Schätzung der Parameter r und b , und somit auch die Schätzung von λ_t sowie die grafische Darstellung in einfacher Art und Weise ermöglicht (Jombart et al. (2019)). Tiefere Informationen zur Poisson Regression und zur Berechnung der Schätzer von r und b können zum Beispiel in Fahrmeir et al. (2009) nachgelesen werden.

Als **Verdopplungszeit** bezeichnen wir jene Zeit in der sich die Zahl der inzidenten Fälle verdoppelt. Ausgehend von $E(y_t) = \lambda_t$ und Formel (1) können wir die Verdopplungszeit als Funktion von r berechnen. Um dies einzusehen sei T die unbekannte Verdopplungszeit und sei y_1 die Anzahl der inzidenten Fälle zu einem beliebigen Zeitpunkt t_1 . Dann suchen wir y_2 zu einem Zeitpunkt t_2 sodass $y_2 = 2y_1$ gilt. Dann bekommen wir $T = t_2 - t_1$ und wir rechnen

$$y_2/y_1 = 2 \Leftrightarrow e^{rt_2+b}/e^{rt_1+b} = 2 \Leftrightarrow e^{r(t_2-t_1)} = 2 \Leftrightarrow T = \ln(2)/r.$$

Effektive Reproduktionszahl

Die effektive Reproduktionszahl, R_{eff} , schätzen wir gemäß dem Ansatz von Cori et al. (2013). Im wesentlichen gehen zwei Parameter des Ausbruchs bzw. der Epidemie in die Schätzung ein. Erstens die Zeitreihe der inzidenten Fälle und zweitens die Verteilung des seriellen Intervalls (Richter et al. (2020a)). Wie sich der jeweilige Einfluss genau darstellt und wie R_{eff} im Detail geschätzt werden kann, wird für die etwas geübteren Leser*innen im folgenden erklärt.

Sei y_t wie oben die Anzahl der neu aufgetretenen (=inzidenten) Erkrankungsfälle an Tag t . Es gilt die gleiche Annahme wie oben, dass y_t Poisson verteilt ist mit Parameter λ_t . Sei w_s die Wahrscheinlichkeit dass ein Infizierter eine weitere Infektion s Tage nach seiner eigenen Infektion verursacht (s wird auch *generation time* genannt). Da die *generation time* meistens unbekannt ist verwendet man in der Regel die Verteilung des seriellen Intervalls als Approximation für w_s . Nehmen wir nun an, dass das serielle Intervall gammaverteilt ist (Richter et al. (2020a)) und die Dichte

$$p(s; k, \theta) := \frac{s^{k-1} e^{-s/\theta}}{\theta^k \Gamma(k)}, s > 0, k > 0, \theta > 0$$

besitzt. Dann wird w_s wie folgt berechnet: Sei $s = 1, \dots, m$, dann ist $w_s = p(s; k, \theta) / \sum_{i=1}^m p(i; k, \theta)$ und es gilt daher $\sum_{i=1}^m w_s = 1$. Die effektive Reproduktionszahl, R_{eff} , sei die durchschnittliche Anzahl an unmittelbaren Folgefällen, generiert pro Infiziertem. Nun gehen wir davon aus, dass die Infizierten zu einem Zeitpunkt $t - s$ mit der Rate $R_{eff} \times w_s$ zur Anzahl der Infizierten am Tag t beitragen. Natürlich leisten aber auch die Infizierten an den anderen Tage, $\{1, \dots, t - s - 1, t - s + 1, \dots, t - 1\}$, einen Beitrag. Da λ_t den mittleren Infizierten am Tag t entspricht gilt somit $\lambda_t = R_{eff} \sum_{s=1}^t y_{t-s} w_s$. Die effektive Reproduktionszahl ist abhängig vom Tag t für den die Zahl geschätzt werden soll, und wir können auch einen Zeitraum τ (z.B. $\tau = 13$ Tage) definieren, die in die Schätzung vom jeweiligen R_{eff} eingehen. Um diese Abhängigkeit von t und τ darzustellen gehen wir zu einer neuen Notation über und bezeichnen die effektive Reproduktionszahl im Folgenden mit $R_{t,\tau}$. Wir nehmen an, dass $R_{t,\tau}$ für die Zeitperiode $[t - \tau + 1, t]$ konstant ist.

Im Folgenden schätzen wir $R_{t,\tau}$ mit dem Ansatz der Bayes-Inferenz (eine einfache Einführung findet sich in Fahrmeir et al. (2009) p.478ff).

Dazu benötigen wir eine Annahme über die *a-priori* Verteilung von $R_{t,\tau}$. Wir wählen dazu eine Gammaverteilung mit Formparameter a und Skalenparameter b . Die zugehörige Wahrscheinlichkeitsdichtefunktion ist gegeben durch

$$p(R_{t,\tau}; a, b) = p(R_{t,\tau}) := \frac{R_{t,\tau}^{a-1} e^{-R_{t,\tau}/b}}{b^a \Gamma(a)}, R_{t,\tau} > 0, a > 0, b > 0, \quad (2)$$

wobei $\Gamma(\cdot)$ die vollständige Gammafunktion ist. Für natürliche Zahlen n gilt $\Gamma(n) = (n - 1)!$.

Dann benötigen wir noch die Likelihoodfunktion der Poissonverteilung von \mathbf{y} , $p(\mathbf{y}; R_{t,\tau})$, welche gegeben ist durch

$$p(\mathbf{y}; R_{t,\tau}) = \prod_{i=t-\tau+1}^t (R_{t,\tau} d_i)^{y_i} \exp(-R_{t,\tau} d_i) \frac{1}{y_i!}, \quad (3)$$

wobei $d_i := \sum_{s=1}^i y_{i-s} w_s$ ist. Die *a-posteriori* Verteilungsdichte von $R_{t,\tau}$ ist dann

$$p(R_{t,\tau}; \mathbf{y}) = \frac{p(\mathbf{y}; R_{t,\tau}) p(R_{t,\tau})}{\int p(\mathbf{y}; R_{t,\tau}) p(R_{t,\tau}) d R_{t,\tau}}. \quad (4)$$

Der Faktor $c = \int p(\mathbf{y}; R_{t,\tau}) p(R_{t,\tau}) dR_{t,\tau}$ wird auch *Normierungsfaktor* genannt.

Setzen wir nun (2) und (3) in (4) ein und vereinfachen die Formel so erhalten wir

$$p(R_{t,\tau}; \mathbf{y}) = R_{t,\tau}^{a + \sum_{i=t-\tau+1}^t y_i - 1} \exp\left(-R_{t,\tau} \left(\frac{1}{b} + \sum_{i=t-\tau+1}^t d_i\right)\right) k(\mathbf{y}, t, a, b),$$

wobei

$$k(\mathbf{y}, t, a, b) = \prod_{i=t-\tau+1}^t \frac{d_i^{y_i} c}{y_i! b^a \Gamma(a)}$$

ein von $R_{t,\tau}$ unabhängiger Term ist. $p(R_{t,\tau}; \mathbf{y})$ ist also proportional zu

$$p(R_{t,\tau}; \mathbf{y}) \propto R_{t,\tau}^{a + \sum_{i=t-\tau+1}^t y_i - 1} \exp\left(-R_{t,\tau} \left(\frac{1}{b} + \sum_{i=t-\tau+1}^t d_i\right)\right) \frac{\frac{1}{b} + \sum_{i=t-\tau+1}^t d_i}{\Gamma(a + \sum_{i=t-\tau+1}^t y_i)},$$

und dementsprechend ist $R_{t,\tau}$ gammaverteilt:

$$R_{t,\tau} \sim \text{Gamma}\left(a + \sum_{i=t-\tau+1}^t y_i, \frac{1}{\frac{1}{b} + \sum_{i=t-\tau+1}^t d_i}\right). \quad (5)$$

Der Mittelwert und somit unser Schätzer für die effektive Reproduktionszahl zum Zeitpunkt t und basierend auf τ Tagen ist

$$\begin{aligned} \hat{R}_{t,\tau} &= \frac{a + \sum_{i=t-\tau+1}^t y_i}{\frac{1}{b} + \sum_{i=t-\tau+1}^t d_i} \\ &= \frac{a + \sum_{i=t-\tau+1}^t y_i}{\frac{1}{b} + \sum_{i=t-\tau+1}^t \sum_{s=1}^i y_{i-s} w_s}. \end{aligned}$$

Ein 95% Konfidenzintervall für $\hat{R}_{t,\tau}$ kann mit dem 2.5% und dem 97.5% Perzentil der Verteilungsfunktion von (5) berechnet werden. Die hier präsentierte Methode zur Schätzung von $R_{t,\tau}$ ist im *R*-Paket *EpiEstim* umgesetzt (Cori et al. (2013)).

Ohne weiteres Vorwissen über $R_{t,\tau}$ werden für die a-priori Gamma-Verteilung üblicherweise die Parameter $a = 1$ und $b = 5$ angenommen.

Eine Anwendung der oben beschriebenen Methodik findet sich beispielsweise in der Publikation von Richter et al. (2020b) .

Referenzen

Cori, A., Ferguson, N.M., Fraser, C., Cauchemez, S., 2013. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *Am J Epidemiol* 178, 1505–1512. <https://doi.org/>

10.1093/aje/kwt133

Fahrmeir, L., Kneib, T., Lang, S., 2009. Regression: Modelle, Methoden und Anwendungen, 2. ed. Springer Berlin Heidelberg.

Jombart, T., Kamvar, Z.N., Cai, J., Pulliam, J., Chisholm, S., FitzJohn, R., Schumacher, J., Bhatia, S., 2019. Reconhub/incidence 1.5. <https://doi.org/10.5281/zenodo.2540217>

Richter, L., Schmid, D., Chakeri, A., Maritschnik, S., Pfeiffer, S., Stadlober, E., 2020a. Schätzung des seriellen Intervalles von COVID19, Österreich [WWW Document]. AGES Wissen Aktuell. URL https://www.ages.at/download/0/0/068cb5fb9f2256d267e1a3dc8d464623760fcc30/fileadmin/AGES2015/Wissen-Aktuell/COVID19/Sch%C3%A4tzung_des_seriellen_Intervalles_von_COVID19_2020-04-08.pdf (accessed 9.4.2020).

Richter, L., Schmid, D., Chakeri, A., Maritschnik, S., Pfeiffer, S., Stadlober, E., 2020b. Epidemiologische Parameter des COVID19 Ausbruchs - Update 08.04.2020, Österreich, 2020 [WWW Document]. AGES Wissen Aktuell. URL https://www.ages.at/download/0/0/86de1e8a4d72a143280eca00fd5abc0a969d87c0/fileadmin/AGES2015/Wissen-Aktuell/COVID19/Update_Epidemiologische_Parameter_des_COVID19_Ausbruchs_2020-04-09.pdf (accessed 17.4.2020).