

Vergleich verschiedener Definitionen für die Reproduktionszahl

Februar 2021
v0.1

Abstract

Die Reproduktionszahl kann auf verschiedene Wege hergeleitet werden. Je nach Definition ergeben sich aber unterschiedlich lange Halbierungsintervalle. Um diese Unterschiede zu untersuchen, werden analytische Lösungen für zwei bekannte Herleitungen beschrieben und die Auswirkungen auf Vervielfachungsintervalle verglichen. Abschließend wird dargestellt, wie die R-Werte ineinander überführt werden können.

1 Übersicht

Motivation In der öffentlichen Debatte um die Covid19-Pandemie ist die Reproduktionszahl eine entscheidende Kenngröße. Einerseits beschreibt die effektive Reproduktionszahl R_{eff} , ob sich die Pandemie ausbreitet oder eher kleiner wird und wie schnell das jeweils geschieht. Andererseits hängt von R_0 , der "natürlichen Reproduktionszahl" (Basisreproduktionszahl) ab, wie umfassend Gegenmaßnahmen ausfallen müssen, um eine exponentielle Ausbreitung des Virus zu verhindern.

Allerdings können die Reproduktionszahlen auf viele Wege definiert werden. Abhängig von der Definition ergeben sich aber unterschiedliche Verdopplungs- bzw. Halbierungsintervalle. Analog können so auch unterschiedliche Reproduktionszahlen aus dem gleichen Datensatz hergeleitet werden.

Einschränkungen Hier wird keine Mathematik betrieben, die nicht schon hunderte zuvor durchgeführt hätten. Beispielhaft sei auf [1] verwiesen, wo schon verschiedene Definitionen der Reproduktionszahl und der Einfluss auf Verdopplungs- und Halbierungsintervalle beschrieben wird.

Ziele Wir wollen uns hier nur mit zwei Definitionen befassen: Die erste erschließt sich direkt aus dem Standard-SIR-Modell (zum Beispiel [2]), die zweite findet man beispielsweise in [3]. Dort wurden auch bereits beide Definitionen sehr umfassend verglichen, insbesondere mit Blick auf die Auswertung realer Daten.

Wir interessieren uns dagegen für die eher theoretische Annahme einer (zumindest stückweise) **konstanten** Reproduktionszahl und das daraus folgende exponentielle Wachstum. Damit sollen folgende Fragen geklärt werden:

1. Welche Interpretation der Reproduktionszahl liegen beiden Definitionen zu Grunde?
2. Wie lange sind Verdopplungs- und Halbierungsintervalle für eine gegebene Reproduktionszahl bei beiden Definitionen?
3. Wann sind die Ergebnisse ähnlich?
4. Wie lassen sich beide Definitionen ineinander überführen?

Gliederung Um diese Fragen zu beantworten, gliedert sich das Dokument in vier weitere Teile. In je einem Abschnitt werden wir die Reproduktionszahl herleiten und die Herleitung interpretieren. Im dritten Abschnitt werden wir die Vergleiche zwischen den Definitionen durchführen. Zum Schluss fassen wir die Ergebnisse noch kurz zusammen.

2 Herleitung über das SIR-Modell

Modell Es wird das einfachste SIR-Modell verwendet, wie es zum Beispiel in [2] beschrieben wird. Mit konstanten Parametern ist es über dieses System von Differentialgleichungen definiert:

$$\frac{d}{dt}S(t) = -\beta \cdot S(t) \cdot I(t) \quad (1)$$

$$\frac{d}{dt}I(t) = \beta \cdot S(t) \cdot I(t) - \nu I(t) \quad (2)$$

$$\frac{d}{dt}R(t) = \nu I(t) \quad (3)$$

$S(t)$ beschreibt die zum Zeitpunkt t noch nicht infizierte Bevölkerung, $I(t)$ den Anteil der aktuell infektiösen Bevölkerung. $R(t)$ den Teil, der die infektiöse Phase bereits hinter sich hat. Zu jedem Zeitpunkt gilt: $S(t) + I(t) + R(t) = 1$. β beschreibt dabei die effektive Kontaktrate, für die gilt: $\beta = \tau \cdot c$. τ beschreibt die Transmissibilität, also die Wahrscheinlichkeit dass ein Infizierter einen einzelnen Kontakt ansteckt. c beschreibt die durchschnittliche Kontaktrate, $\nu^{-1} = d$ beschreibt die mittlere Dauer der infektiösen Periode.

Definition der Reproduktionszahl In [2] wird die Basisreproduktionszahl R_0 definiert als: $R_0 = \tau \cdot c \cdot d = \beta \cdot d$. Die effektive Reproduktionszahl $R_{eff} =: R$ ist abhängig von R_0 , der Zahl der bisher Infizierten und eventuellen Verhaltensänderungen.

Interpretation Aus dieser Definition kann man direkt ablesen, welche Faktoren die Reproduktionszahl beeinflussen - also auch an welchen Stellen angegriffen werden kann, um sie zu senken. Zum Beispiel kann die Kontaktrate c durch Bewegungseinschränkungen reduziert werden, die Transmissibilität τ durch Masken oder Verlagerung der Kontakte nach draußen oder die "effektive" Dauer der Infektiosität durch Quarantäne und Isolation (letzteres könnte man natürlich auch als Einfluss auf die Kontaktrate sehen).

Außerdem wird direkt ersichtlich, dass eine (verhältnismäßig) kleine Reduktion an zwei Stellschrauben eine relativ große Wirkung auf das Gesamtergebnis hat, weil sich die Effekte multiplizieren.

Exponentielles Wachstum im SIR-Modell Dieses Modell ist letztlich das Paradebeispiel, für exponentielle Ausbreitung von Krankheiten, solange die Bevölkerung ihr Verhalten nicht ändert. Nimmt man an, dass der größte Teil der Bevölkerung nicht infiziert ist, kann man schreiben $S(t) \approx 1$.

Damit gilt für (2):

$$\frac{d}{dt}I(t) \approx \beta \cdot I(t) - \nu \cdot I(t) = (\beta - \nu) \cdot I(t)$$

Diese Differentialgleichung lässt sich offensichtlich geschlossen lösen zu:

$$I(t) = e^{(\beta - \nu) \cdot t} \quad (4)$$

Die Zahl der Neuinfektionen ergibt sich nun aus der (negativen) Änderung am *Susceptible*-Teil der Differentialgleichungen. Also:

$$\dot{I}_{neu,1}(t) = -\frac{d}{dt}S(t) = \beta \cdot I(t) = \beta e^{(\beta - \nu) \cdot t} \quad (5)$$

Tatsächlich haben wir durch diesen Vorfaktor nichts gewonnen. Nehmen wir zum Beispiel die Startbedingung $S(t=0) = 1.0 - 10^{-6}$, $I(t=0) = 10^{-6}$ und $R(t=0) = 0$ an, dann ergibt es intuitiv Sinn zu fordern, dass gelten muss $\dot{I}_{neu}(t=0) = 10^{-6}$, wenn wir unterstellen dass zuvor keine Infektionen vorhanden waren.

In diesem Fall müssen wir dann also schreiben:

$$\dot{I}_{neu,1}(t) = 10^{-6} \cdot e^{(\beta - \nu) \cdot t}$$

Falls also die Zahl der Neuinfektionen I_0 zum Zeitpunkt $t=0$ bekannt ist, können wir schreiben:

$$\dot{I}_{neu,1}(t) = I_0 \cdot e^{(\beta - \nu) \cdot t} \quad (6)$$

Zusammenhang zwischen exponentiellem Wachstum und der Reproduktionszahl Da wir letztlich an einem Vergleich der verschiedenen Definitionen interessiert sind, suchen wir einen Ausdruck der Form:

$$\dot{I}_{neu,i}(t) = I_0 \cdot e^{r_i(R) \cdot t} \quad (7)$$

$r_i(R)$ ist dabei ein von der Reproduktionszahl abhängiger Faktor, der die Dauer der Halbierungs- bzw. Verdopplungszeiten bestimmt. In diesem Fall erhalten wir:

$$r_1(R) = \frac{R-1}{d} \quad (8)$$

Es gibt offensichtlich einen linearen Zusammenhang zwischen r_1 und R .

3 Herleitung über Fallzahlen-Definition

Definition der Reproduktionszahl Aus [3] kennen wir eine andere triviale Definition für die Reproduktionszahl für zu diskreten Zeitpunkten angegebene Neuinfektionen:

$$R_t = \frac{C_t}{C_{t-g}} \quad (9)$$

Mit g ist hier die Generationszeit gemeint, wir werden aber annehmen, dass $g = d$. C_t gibt die Neuinfektionen zum Zeitpunkt t an, C_{t-g} die Neuinfektionen zum Zeitpunkt $t - g$. In diesem Fall muss g also erstmal eine positive Ganzzahl sein.

Interpretation Diese Definition der Reproduktionszahl ist weniger ergiebig in ihrer Bedeutung. Im Prinzip wird hier vereinfachend angenommen, dass eine vor g Tagen infizierte Person heute R_t Menschen ansteckt. Das entspricht ziemlich genau der "landläufig bekannten" Definition: "Die Reproduktionszahl gibt an, wie viele Menschen eine infizierte Person im Durchschnitt ansteckt."

Somit ist die große Stärke dieser Sichtweise ihre intuitive Verständlichkeit, dass hier exponentielles Wachstum stattfindet: Steckt eine infizierte Person immer zwei Leute an, wird klar wie schnell eine unglaublich große Zahl an Infizierten entstehen kann.

Eine weitere Stärke dieser Definition: Die Reproduktionszahl kann hier schon zeitabhängig sein. Da wir ja aber konstante Parameter unterstellen, ignorieren wir das im Folgenden.

Zusammenhang zwischen exponentiellen Wachstum und der Reproduktionszahl Die Herleitung des exponentiellen Wachstums ist hier recht unelegant. Weil wir ja letztlich eine Funktion suchen, schreiben wir statt C_t lieber $c[k]$. k ist hier einfach unser Zeitindex, um deutlich zu machen, dass wir hier diskrete Zeitschritte haben. Somit muss $c[k]$ erfüllen:

$$c[k] = R \cdot c[k - d]$$

Eine Lösung für diese Aufgabe ist gegeben durch:

$$c[k] = R^{k/d} \quad (10)$$

Das können wir einfach durch Einsetzen überprüfen:

$$R \cdot c[k - d] = R \cdot R^{(k-d)/d} = R \cdot R^{k/d - d/d} = R \cdot R^{k/d} \cdot R^{-1} = R^{k/d} = c[k]$$

Erweiterung zum kontinuierlichen Fall Analog zum Abschnitt 2. interessieren wir uns auch hier für eine Definition mit kontinuierlichen Zeitschritten, schon allein wegen der besseren Vergleichbarkeit aber auch um nicht weiter fordern zu müssen, dass die Generationszeit eine Ganzzahl ist.

Jetzt fordern wir also:

$$c(t) = R \cdot c(t - d) \quad (11)$$

Auch hier ist die Lösung analog zum diskreten Fall:

$$c(t) = R^{t/d} \quad (12)$$

Wir überprüfen wieder durch einsetzen:

$$R \cdot R^{(t-d)/d} = R \cdot R^{(t/d-d/d)} = R \cdot R^{t/d} \cdot R^{-1} = R^{t/d} = c(t)$$

Um die Form $\dot{I}_{neu,i}(t) = e^{r_i(R) \cdot t}$ zu erhalten, verwenden wir den Zusammenhang $a^x = e^{x \cdot \ln(a)}$. Somit erhalten wir:

$$\dot{I}_{neu,2}(t) := I_0 \cdot c(t) = I_0 \cdot e^{\ln(R)/d \cdot t} \quad (13)$$

Für $r_2(R)$ erhalten wir damit:

$$r_2(R) = \frac{\ln(R)}{d} \quad (14)$$

4 Vergleich der beiden Definitionen

Rekapitulation Bis jetzt wissen wir also, dass beide Herleitungen für konstante Parameter auf ein exponentielles Wachstum führen. Wir erhalten also eine Funktion der Form:

$$\dot{I}_{neu,i}(t) = I_0 \cdot e^{r_i(R) \cdot t}$$

Für die Herleitung über das SIR-Modell gilt:

$$r_1(R) = \frac{R-1}{d}$$

Definieren wir die Reproduktionszahl über die Neuinfektionen, erhalten wir:

$$r_2(R) = \frac{\ln(R)}{d}$$

Gemeinsame Eigenschaften Beide Definitionen stellen sicher, dass grundlegende Eigenschaften des R-Wertes erhalten bleiben:

1. Für $R = 1$ wird $r_i(R) = 0$, die Zahl der täglichen Neuinfektionen bleibt also konstant.
2. Für $R > 1$ wird $r_i(R) > 0$, die Zahl der täglichen Neuinfektionen steigt also mit fortlaufendem t . Je größer R , desto schneller steigt die Zahl der täglichen Neuinfektionen.
3. Für $R < 1$ wird $r_i(R) < 0$, die Zahl der täglichen Neuinfektionen fällt also. Je näher R sich an die Null annähert, desto schneller sinkt die Zahl der täglichen Neuinfektionen.

Insofern sind beide Definitionen also recht ähnlich, als dass die zentralen Eigenschaften des R-Wertes beibehalten werden.

Interpretation als Taylorpolynom Tatsächlich lässt sich ein sehr eleganter Zusammenhang zwischen $r_1(R)$ und $r_2(R)$ finden: $r_1(R)$ kann als ein Taylor-Polynom erster Ordnung zu $r_2(R)$ um $R = 1$ gesehen werden, es gilt also:

$$r_1(R) = T_{r_2,1,1}(R) = r_2(1) \cdot (R-1)^0 + r_2'(1) \cdot (R-1)$$

Mit $r_2'(R) = \frac{1}{R \cdot d}$, $r_2(1) = 0$ und $r_2'(1) = \frac{1}{d}$ folgt dann:

$$r_1(R) = 0 + \frac{1}{d} \cdot (R-1) = \frac{R-1}{d}$$

Fehlerabschätzung mit Taylor Mit dieser Definition können wir die Abschätzung für die Taylor-Entwicklung nutzen, um den Unterschied zwischen $r_1(R)$ und $r_2(R)$ nach oben zu begrenzen.

Allgemein gilt für ein Taylorpolynom $T_{f,x_*,n}$ der n -ten Ordnung für eine Funktion f um x_* :

$$|f(x) - T_{f,x_*,n}(x)| \leq \frac{C}{(n+1)!} |x - x_*|^{n+1}$$

C beschränkt dabei $|f^{(n+1)}(x)|$ zwischen x und x_* nach oben. Auf unser Problem bezogen gilt also:

$$|r_2(R) - r_1(R)| \leq \frac{C}{2!} |R-1|^2$$

Um C zu erhalten, suchen wir das Maximum von $r_2^{(2)}(R)$ zwischen R und 1. Die zweite Ableitung erfüllt: $r_2^{(2)}(R) = -\frac{1}{d \cdot R^2}$. Wir unterscheiden zwei Fälle, um eine möglichst optimale Fehlerabschätzung zu erlauben:

1. Fall 1: $R \geq 1$

In diesem Fall wird $r_2^{(2)}$ maximal bei $R = 1$. Somit gilt für C : $C = \frac{1}{d}$.

2. Fall 2: $R \leq 1$

Hier lässt sich leider keine feste Obergrenze finden, da $r_2^{(2)}(R)$ für $R \rightarrow 0$ gegen $-\infty$ strebt. Stattdessen liegt die kleinste Obergrenze beim jeweiligen R , somit gilt hier: $C = \frac{1}{d \cdot R^2}$.

Praktische Konsequenzen aus Taylor Dieser Zusammenhang zwischen $r_1(R)$ und $r_2(R)$ garantiert uns, dass sich beide Definitionen nur geringfügig unterscheiden, wenn R nicht zu weit von 1 abweicht.

An dieser Stelle noch ein exemplarisches Beispiel für den Fall $R = 0.6$. $r_1(R)$ erfüllt (mit $d=4$): $r_1(0.6) = -0.100$. Für $r_2(R)$ gilt dagegen: $r_2(0.6) \approx -0.128$. Der Unterschied erscheint auf den ersten Blick gering, aber dabei ist zu beachten, dass $r_i(R)$ im Exponenten steht. Kleine Änderungen können also große Unterschiede verursachen.

Halbierungsintervalle Um anschaulich zu quantifizieren, wie schnell sich Fallzahlen für ein gegebenes R verringern, wird häufig angegeben wie lange eine Halbierung dauert. Das heißt wir fragen, für welches $\Delta_{H,i} > 0$ gilt:

$$\dot{I}_{neu,i}(t + \Delta_{H,i}) = \frac{1}{2} \dot{I}_{neu,i}(t) \quad (15)$$

Um die Frage etwas einfacher zu beantworten, setzen wir $t = 0$, was keine Einschränkung darstellt.

$$\dot{I}_{neu,i}(0 + \Delta_{H,i}) = \frac{1}{2} \dot{I}_{neu,i}(0)$$

Setzen wir für $\dot{I}_{neu,i}(\Delta_{H,i})$ ein, wird die Frage deutlich einfacher, wenn wir auf beiden Seiten I_0 kürzen:

$$e^{r_i(R) \cdot \Delta_{H,i}} = \frac{1}{2} \cdot e^{r_i(R) \cdot 0} = \frac{1}{2}$$

Somit können wir einfach umstellen, um die Dauer eines Halbierungsintervalls zu erhalten:

$$\Delta_{H,i} = \frac{\ln(\frac{1}{2})}{r_i(R)} \quad (16)$$

Für die Herleitung aus dem SIR-Modell erhalten wir:

$$\Delta_{H,1} = \ln(\frac{1}{2}) \cdot \frac{d}{R-1}$$

Über die Fallzahlen-Definition ergibt sich:

$$\Delta_{H,2} = \ln(\frac{1}{2}) \cdot \frac{d}{\ln(R)}$$

Aus diesen beiden Definitionen ergibt sich ein ganz praktischer Unterschied: Wird die Reproduktionszahl über das SIR-Modell definiert, kann die Halbierung nicht beliebig schnell gehen ($\Delta_{H,1}$ wird nie 0). $\Delta_{H,2}$ strebt zumindest gegen 0, wenn $R \rightarrow 0$.

Ausgewählte Halbierungsintervalle

R	Δ_1	Δ_2
1	∞	∞
0.95	55	54
0.9	28	26
0.8	14	12
0.7	9	8
0.6	7	5
0.5	6	4
0.4	5	3
0.1	3 (3.08)	1 (1.20)
0.05	3 (2.92)	1 (0.93)
0.01	3 (2.80)	1 (0.60)
0.001	3 (2.77)	1 (0.40)

An diesen Beispielwerten ($d = 4$) sieht man deutlich, dass die Herleitung über die Fallzahlen zu kürzeren Halbierungsintervallen führt, als die Herleitung über das SIR-Modell. Außerdem sieht erkennt man bei den sehr niedrigen Reproduktionszahlen, wie $\Delta_{H,2}$ gegen 0 strebt, $\Delta_{H,1}$ aber nicht.

Die Unterschiede machen deutlich, dass beide Definitionen nicht willkürlich ausgetauscht werden können und hier tatsächlich Vorsicht geboten ist. Insbesondere um $R \approx 0.7$ macht sich auch ein kleiner psychologischer Unterschied recht deutlich bemerkbar: Ein Halbierungsintervall von einer Woche klingt deutlich kürzer als 9 Tage, auch wenn nur zwei Tage Unterschied dazwischen liegen.

Allgemeine Vervielfachungsintervalle Auch wenn Halbierungsintervalle ein relativ leicht verständliches Modell sind, ist es natürlich auch interessant, wie lange es dauert, bis eine beliebige Reduktion (oder auch Erhöhung) der Fallzahlen eingetreten ist. Wird die Inzidenz J_{Ende} angestrebt und liegt aktuell die Inzidenz J_{Start} vor, dann ergibt sich:

$$\Delta_{V,1} = \ln\left(\frac{J_{Ende}}{J_{Start}}\right) \cdot \frac{d}{R-1}$$

beziehungsweise

$$\Delta_{V,2} = \ln\left(\frac{J_{Ende}}{J_{Start}}\right) \cdot \frac{d}{\ln(R)}$$

Benötigter R-Wert für ein bestimmtes Halbierungsintervall Eine andere interessante Frage ist, welcher R-Wert nötig ist, damit ein bestimmtes Halbierungsintervall erreicht wird. Dazu stellen wir einfach Gleichung (16) um:

$$r_i(R) = \frac{\ln(\frac{1}{2})}{\Delta_{H,i}}$$

Ist $r_i^{-1}(\cdot)$ die Umkehrfunktion zu $r_i(R)$, dann gilt:

$$R = r_i^{-1}\left(\frac{\ln(\frac{1}{2})}{\Delta_{H,i}}\right) \quad (17)$$

Mit $r_1^{-1}(y) = d \cdot y + 1$ folgt:

$$R = \ln\left(\frac{1}{2}\right) \cdot \frac{d}{\Delta_{H,1}} + 1$$

Analog erhalten wir für $r_2^{-1}(y) = e^{y \cdot d}$:

$$R = e^{d \cdot \frac{\ln(\frac{1}{2})}{\Delta_{H,2}}}$$

Umrechnung der Definitionen Bis jetzt haben wir implizit immer gefragt welche Konsequenzen sich aus einer bestimmten Reproduktionszahl in den zwei Modellen folgen. Mindestens genauso interessant ist aber, wie wir beide Modelle ineinander umrechnen können. Zur Motivation stellen wir uns folgendes Szenario vor: Wir haben mit der Fallzahlen-Definition die (zeitvariable) Reproduktionszahl geschätzt. Nun wollen wir aber in einem SIR-Modell mit diesen

Ergebnissen weiter rechnen. Dazu wäre es interessant, die ermittelte Reproduktionszahl an das SIR-Modell anzupassen.

Da wir nun mit zwei unterschiedlichen R-Werten rechnen, verwenden wir ab sofort die Notation R_1 und R_2 . Der Ansatz zur Umrechnung lautet:

$$r_1(R_1) = r_2(R_2) \quad (18)$$

$$\frac{R_1 - 1}{d} = \frac{\ln(R_2)}{d}$$

Zuerst lösen wir nach R_1 auf:

$$R_1 = \ln(R_2) + 1 \quad (19)$$

Lösen wir umgekehrt nach R_2 auf, erhalten wir:

$$R_2 = e^{R_1 - 1} \quad (20)$$

5 Zusammenfassung

Bedeutung der Definitionen Das SIR-Modell führt zu einer Herleitung, die zeigt von welchen Faktoren der R-Wert abhängt. Die Definition über die Fallzahlen zeigt dagegen eher die Auswirkungen des R-Werts, nämlich wie viele Personen eine infizierte Person im Mittel ansteckt.

Vergleich Beide Definitionen führen - für konstante Parameter - auf exponentielles Wachstum. Dabei erhalten sie grundsätzliche Eigenschaften, wann sich die Pandemie ausbreitet, wann sie gleich bleibt oder wann sie zurückgedrängt wird.

Solange $R \approx 1$, ist es unproblematisch, beide Definitionen auszutauschen. Für größere Abweichungen zeigen sich recht deutliche Unterschiede in den Vielfachungsintervallen, allerdings gibt es durch den Zusammenhang mittels Taylor eine einfache Möglichkeit zur Fehlerabschätzung.

Schließlich lassen sich beide Definitionen zumindest für konstante Parameter ineinander umrechnen.

References

- [1] Wallinga J and Lipsitch M. “How generation intervals shape the relationship between growth rates and reproductive numbers”. In: *Proc. R. Soc. B.* 274 (2007), pp. 599–604. DOI: <https://doi.org/10.1098/rspb.2006.3754>.
- [2] James Holland Jones. *Notes On R0*. 2007. URL: <https://web.stanford.edu/~jhj1/teachingdocs/Jones-on-R0.pdf>.
- [3] Jonas Dehning et al. “Model-based and model-free characterization of epidemic outbreaks”. In: *medRxiv* (2020). DOI: 10.1101/2020.09.16.20187484. eprint: <https://www.medrxiv.org/content/early/2020/09/18/2020.09.16.20187484.full.pdf>. URL: <https://www.medrxiv.org/content/early/2020/09/18/2020.09.16.20187484>.