



Technische
Universität
Braunschweig



Schätzung der Verteilung der individuellen Reproduktionszahl: Der Fall der COVID-19-Pandemie

Prof. Dr. Jens-Peter Kreiß

Studium Generale an der TU Braunschweig am 26. Oktober 2021

Mit wem haben Sie es zu tun?

Meine Name ist **Jens-Peter Kreiß** und ich bin der Leiter des Instituts für Mathematische Stochastik an der TU Braunschweig.

Stochastik könnte man etwas frei übersetzen mit

im Vermuten geschickt sein.

Das würden wir, und diese Arbeit ist gemeinsam mit meinen Kollegen

- Mag. Alexander Braumann (TU Braunschweig)
 - Dr. Jonas Krampe (Universität Mannheim)
 - Prof. Dr. Efstathios Paparoditis (University of Cyprus, Nicosia),
- heute gerne am Beispiel von Corona unter Beweis stellen.

Überblick

Einführung

Die individuelle Reproduktionszahl R

Zufallsvariablen und wichtige Kenngrößen

Modellierung der Reproduktionszahl R

Modellanpassung durch Parameterschätzung

Modellrechnungen für Wahrscheinlichkeiten für Sekundärinfektionen

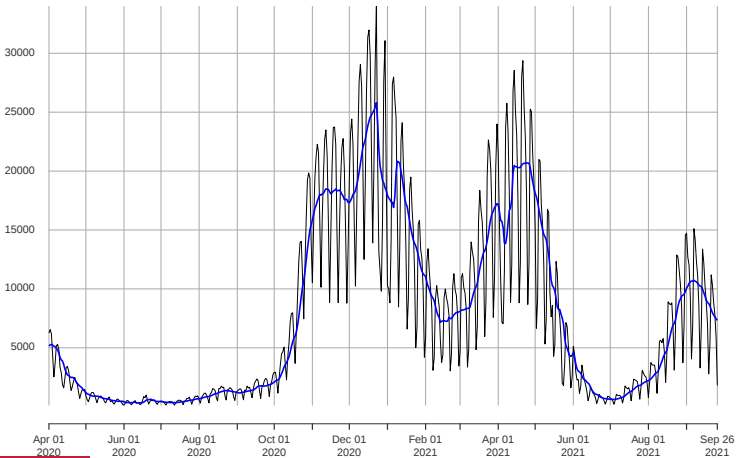
Zusammenfassung

Die Corona-Pandemie

Die Corona-Pandemie hält die ganze Welt seit März 2020 fest im Griff.
Die folgenden Begriffe sind in aller Munde:

- COVID-19
- SARS-CoV-2-Virus
- Lockdown
- AHA+L-Regel
- 7-Tage Inzidenz
- Antigen-Schnelltest und PCR-Test
- Übersterblichkeit
- R -Wert und exponentielles Wachstum
- mRNA Impfstoffe (BioNTech und Moderna) und Vektorimpfstoffe (AstraZeneca und Johnson&Johnson)
- Impfdurchbrüche
- 3G-Regel

Der Verlauf der Infektionszahlen in Deutschland



Der Verlauf der Infektionszahlen in Deutschland

Beim Verlauf der Infektionszahlen in Deutschland fällt folgendes auf:

- deutliche Variation der absoluten Zahlen
- wellenförmiges Verhalten
- bisher massivste Welle im Winter 2020/2021
- starke Fluktuationen über die Woche
- ...

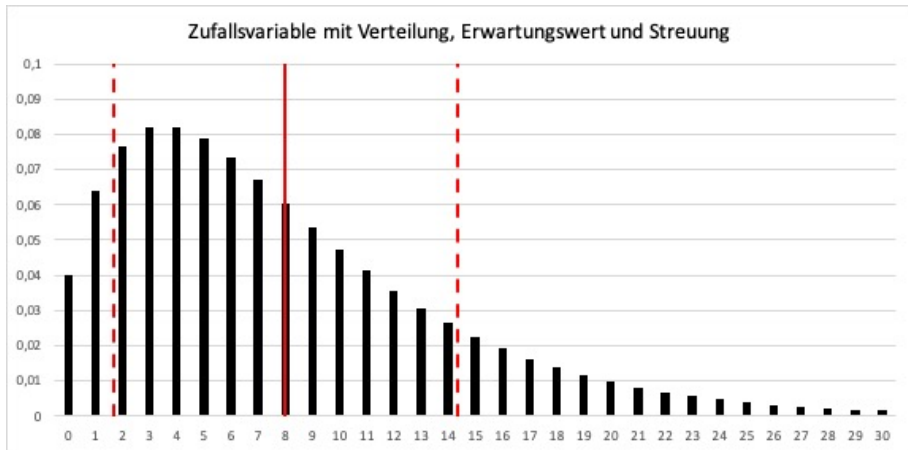
Die individuelle Reproduktionszahl R

Die individuelle Reproduktionszahl R beschreibt die zufällige Anzahl der Ansteckungen, die von einer einzelnen SARS-CoV-2-positive Person ausgelöst werden. Man spricht von sog. Sekundärinfektionen.

R ist eine in der Epidemiologie weit verbreitete Kennzahl, um das Ansteckungspotential einer Infektionserkrankung zu quantifizieren.

Zufallsbeeinflusste Größen (auch *Zufallsvariablen* genannt) besitzen eine sogenannte (Wahrscheinlichkeits-) *Verteilung*.

Zufallsvariablen mit Kenngrößen



Die individuelle Reproduktionszahl R

Die individuelle Reproduktionszahl R beschreibt die zufällige Anzahl der Sekundärinfektionen durch eine einzelne SARS-CoV-2-positive Person.

Zunächst ist der sog. *Erwartungswert* $E(R)$ der Reproduktionszahl, der häufig mit R_0 (auch *Basis-Reproduktionszahl*) bezeichnet wird, von Interesse.

$R_0 < 1$: Indiz dafür, dass die Pandemie unter "Kontrolle" ist

$R_0 > 1$: Warnung für Zustand eines exponentiellen Wachstums

Aber: Der *Erwartungswert* ist **nur eine Kenngröße (Parameter)** der Verteilung der Zufallsgröße R !

Uns interessiert die gesamte **Verteilung** von R mit ihren Wahrscheinlichkeiten dafür, dass eine SARS-CoV-2-positive Person genau k Personen ansteckt ($k = 0, 1, 2, \dots$).

Modellierung der Reproduktionszahl R

Relative Häufigkeiten für die Anzahl k der durch eine SARS-CoV-2 positive Person ausgelösten Sekundärinfektionen können nicht (oder nur durch ein extrem aufwändiges *contact-tracing*) beobachtet werden. Wir benötigen ein **Modell**!

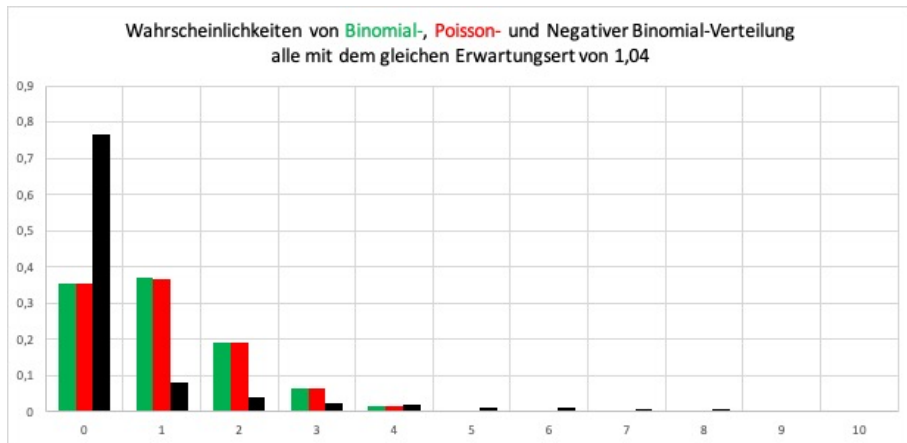
Es kommt die *Binomialverteilung* und die Approximation an diese für seltene Ereignisse, die *Poissonverteilung*, in Frage.
(Vgl. "Das Gesetz der kleinen Zahlen" von L. von Bortkiewicz (1898)).

Ein Blick in die Literatur der Epidemiologie bringt die *Negative Binomialverteilung*, die in ihrer Grundform das Warten auf Erfolge in 0-1-Zufallsketten beschreibt, als alternative Modellverteilung ins Spiel.

Die Negative Binomialverteilung erlaubt im Gegensatz zur Poissonverteilung eine sog. **Überdispersion**!

Modellierung der Reproduktionszahl R

Schauen wir uns die drei Verteilungen im Vergleich an:



Modellierung der Reproduktionszahl R

Während sich die Binomial- und die Poisson-Verteilung nur ganz unwesentlich unterscheiden, erkennen wir zur Negativen Binomialverteilung sehr deutliche Unterschiede.

Hauptunterschied: Die Wahrscheinlichkeit für $k = 0$ Sekundärinfektionen ist bei der Negativen Binomialverteilung deutlich höher.

Ist damit verbunden, dass bei der Negativen Binomialverteilung geringere Wahrscheinlichkeiten für $k \geq 1$ Sekundärinfektionen vorliegen?

Schauen wir genauer hin!

Modellierung der Reproduktionszahl R



Modellierung der Reproduktionszahl R

Die Negative Binomial-Verteilung legt, gegeben, dass überhaupt Ansteckungen aufgetreten sind, offenbar **mehr** Wahrscheinlichkeit auf höhere Anzahlen von Sekundäransteckungen.

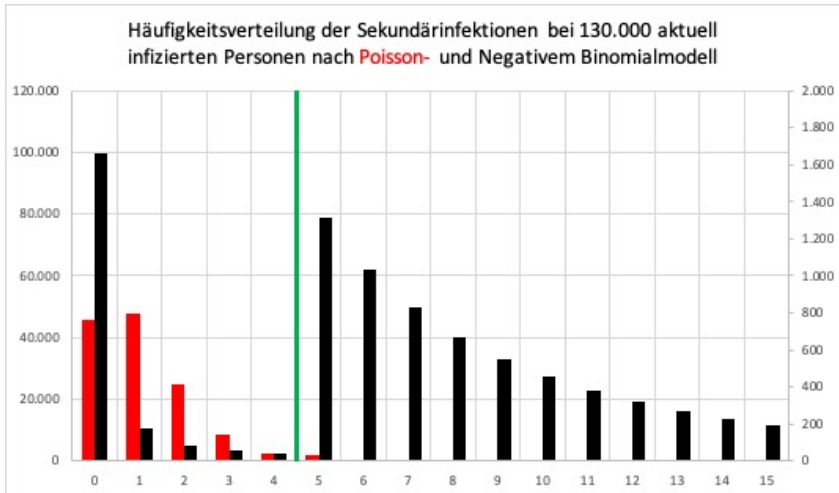
Das gibt einen gewissen Einblick in das, was hinter dem Begriff der sog. **Überdispersion** steht. Anders formuliert: Bei Überdispersion ist die Streuung in den Randbereich großer Anzahlen stärker ausgeprägt.

Vielleicht etwas provokant formuliert:

Die Negative Binomialverteilung ist, im Gegensatz zur Poissonverteilung, in der Lage *Superspreading* zu erfassen

Eine weitere Illustration dazu:

Modellierung der Reproduktionszahl R



Zusammenhang Poisson- und Negative Binomialverteilung

Die Poisson- und die Negative Binomial-Verteilung hängen zusammen.

Beide Verteilungen besitzen Einstellgrößen (Parameter), die eine Anpassung an real beobachtete Daten erlauben.

Die Poisson-Verteilung besitzt nur einen solchen Parameter (λ) und die Negative Binomialverteilung besitzt zwei Parameter (r und p).

Die Poisson-Verteilung geht in die Negative Binomialverteilung über, wenn man annimmt, dass der Parameter λ individuell festgelegt ist, d.h. der Parameter verändert sich über die Personen und besitzt sozusagen selbst eine Verteilung über die Bevölkerung.

Ein solches Resultat korrekt zu beweisen ist für Mathematikinteressierte natürlich sehr spannend! Aber wir wollen uns hier mit den Möglichkeiten der Parameterwahl anhand von zugänglichen realen Corona-Daten auseinandersetzen.

Modellanpassung aus zur Verfügung stehenden Daten

Das Robert-Koch-Institut (RKI), die zentrale Einrichtung der Bundesregierung auf dem Gebiet der Krankheitsüberwachung und -prävention, liefert täglich den sog. 7-Tage R-Wert.

Vielleicht hilft uns dieser 7-Tage R-Wert bei unserem Wunsch nach Anpassung einer Wahrscheinlichkeitsverteilung für zur Beschreibung der individuellen Anzahlen von Sekundärinfektionen.

Der 7-Tage R-Wert des RKI

Aber was gibt der 7-Tage R-Wert des RKI an?

Die Formel zur Berechnung lautet:

$$\hat{R}_{0,t}^7 := \frac{\sum_{s=0}^6 \text{RKI}_{t-s}}{\sum_{s=0}^6 \text{RKI}_{t-4-s}}$$

Diese Formel müssen wir verstehen!

Modellanpassung aus zur Verfügung stehenden Daten

Zur Anpassung einer Verteilung zur Beschreibung der individuellen Sekundärinfektionen benötigen wir eigentlich für jede an einem bestimmten Tag t infizierte Person die Anzahl der jeweiligen Sekundärinfektionen, d.h. die von dieser Person in der Folge angesteckten Personen.

Derartige Daten stehen aber nicht zur Verfügung!

Uns stehen lediglich die Gesamtanzahlen, der an den Tagen t positiv auf SARS-CoV-2 getesteten Personen zur Verfügung!

Helfen uns diese Zahlen überhaupt weiter?

Wir werfen zunächst einen Blick auf die Entstehung dieser Daten durch die sogenannten *Meldekette*.

Die Meldekette

- Positive SARS-CoV-2-Tests müssen dem Gesundheitsamt des jeweiligen Landkreises innerhalb von 24 Stunden gemeldet werden
- Mit etwa 24 Stunden Verzögerung geht die Meldung weiter an das Landesgesundheitsamt und von dort ans RKI
- Am Tag nach der RKI-Meldung werden die Anzahlen veröffentlicht; mit Datum der ersten Meldung
- Symptome treten im Mittel vier Tage nach Infektion auf
- 1-2 Tag bis PCR-Ergebnis vorliegt und 1-2 Tage Meldeverzug

bedeuten in Konsequenz: Das Meldedatum eines positiven Tests liegt in etwa 7 Tage nach dem eigentlichen Infektionsdatum!

Under-Reporting

Das bedeutet: Wir haben von dem Zeitpunkt der Infektion bis zum Meldedatum eine Zeitverzögerung von ungefähr 7 Tagen.

Die vom RKI veröffentlichten Anzahlen geben also (ungefähr) die Gesamtzahl der **vor einer Woche** an SARS-CoV-2 infizierten Personen, genauer, der an SARS-CoV-2 infizierten und **PCR-getesteten** Personen wieder.

D.h., dass das RKI nicht alle infizierten Personen erfasst! Es liegt ein sogenanntes *under-reporting* vor!

Und von diesen Personen beobachten wir dann erneut eine Woche später die Gesamtanzahl der aktenkundig gewordenen Sekundärinfektionen.

Modellanpassung aus zur Verfügung stehenden Daten

Modellanpassung aus zur Verfügung stehenden Daten

Damit können wir **keine individuellen Sekundärinfektionen** beobachten, sondern lediglich aufsummierte Sekundärinfektionen über eine große Anzahl von primär infizierten Personen.

Sekundärinfektionen treten ca. 4 Tage nach der Primärinfektion auf (Generationszeit des Virus).

Daraus können wir über den Quotienten aber auf eine mittlere Anzahl von Sekundärinfektionen pro primär infizierter Person schließen; den **Erwartungswert**.

Für die Poissonverteilung reicht dies bereits aus, um den einen vorhandenen Parameter λ anzupassen!

Für die Negative Binomialverteilung reicht dies aber nicht aus!

Der 7-Tage R-Wert des RKI

Wir schauen noch etwas genauer auf die Berechnungsmethode des RKI:

Das RKI dividiert nicht direkt die aktuell veröffentlichte Anzahl der positiven Tests mit dem Wert, der vor 7 Tagen veröffentlicht wurde.

Das RKI mittelt beide Werte jeweils noch über eine Woche aus, um Schwankungen, die auf ein unterschiedliches Meldeverhalten an einzelnen Wochentage zurückzuführen sind, auszugleichen. Das führt zu:

$$\begin{aligned}\hat{R}_{0,t}^7 &:= \frac{\sum_{s=0}^6 \text{RKI}_{t-s}}{\sum_{s=0}^6 \text{RKI}_{t-4-s}} = \frac{\text{RKI}_t + \text{RKI}_{t-1} + \dots + \text{RKI}_{t-6}}{\text{RKI}_{t-4} + \text{RKI}_{t-5} + \dots + \text{RKI}_{t-10}} \\ &\approx \frac{\text{RKI}_{t-3}}{\text{RKI}_{t-7}} \approx \frac{\text{gemeldete Neuinfektionen am Tag } t-10}{\text{gemeldete Neuinfektionen am Tag } t-14}.\end{aligned}$$

Der 7-Tage R-Wert des RKI

Der 7-Tage R-Wert des RKI

$$\hat{R}_{0,t}^7 \approx \frac{RKI_{t-3}}{RKI_{t-7}} \approx \frac{\text{gemeldete Neuinfektionen am Tag } t-10}{\text{gemeldete Neuinfektionen am Tag } t-14}.$$

beschreibt approximativ die mittlere Anzahl an gemeldeten Sekundärinfektionen von gemeldeten Primärinfektionen, die vor 14 Tagen stattgefunden haben!

Wir haben damit insgesamt einen Zeitverzug von 14 Tagen!!

Aber was ist mit dem **under-reporting** durch Infektionen, die nicht zu einem PCR-Test und damit nicht zu einer Meldung geführt haben?

Bereitet uns das Probleme?

Under-Reporting

Infektionen, die nicht zu einem PCR-Test geführt haben, können vom RKI nicht berücksichtigt werden. Ein Grund für ein under-reporting sind asymptomatische Infektionsverläufe. Es wird geschätzt, dass etwas 20 – 45% der Infektionen asymptomatisch verlaufen.

Aber vielleicht führen auch sehr leichte Infektionsverläufe nicht zu PCR-Tests!

Auch hierzu gibt es Studien. So etwa durch Reihentestungen in dem Ort Kupferzell nach einem Konzert mit vielen Ansteckungen.

Die **reporting-Raten** belaufen sich nach diesen Studien zwischen **nur 20%** bis zu maximal 50%.

D.h. wir müssen mit einem **under-reporting von bis zu 80%** rechnen!

Auswirkungen des Under-Reporting

Zurück zu unserer Ausgangsfrage: Bereitet das **under-reporting** bei der Verwendung des 7-Tage R-Wertes als Schätzer für den Erwartungswert der Anzahl der individuellen Sekundärinfektionen **Probleme**?

Geht man davon aus, dass sich der **Anteil** der **gemeldeten** SARS-CoV-2 infizierten Personen nicht zu schnell und nicht zu stark über die Zeit verändert, so kürzt sich der im Zähler und im Nenner der Formel des RKI auftretende gleiche Anteilsfaktor heraus! D.h.:

$$\hat{R}_{0,t}^7 \approx \frac{RKI_{t-3}}{RKI_{t-7}} \approx \frac{\text{Neuinfektionen am Tag } t-10}{\text{Neuinfektionen am Tag } t-14}.$$

Modellanpassung aus zur Verfügung stehenden Daten

Wir halten fest:

- Der 7-Tage R-Wert des RKI ist eine Schätzung für den Erwartungswert der individuellen und zufälligen Anzahl angesteckter Personen!
- Der 7-Tage R-Wert des RKI gibt das Ansteckungsverhalten von vor etwa 14 Tagen wieder!
- Hinter der Berechnung des 7-Tage R-Wertes stehen Annahmen!

Etwa: Auftreten von Symptomen und Testverhalten ändern sich nur langsam über die Zeit

Ansteckungsverhalten von gemeldeten und nicht gemeldeten infizierten Personen ist ähnlich.

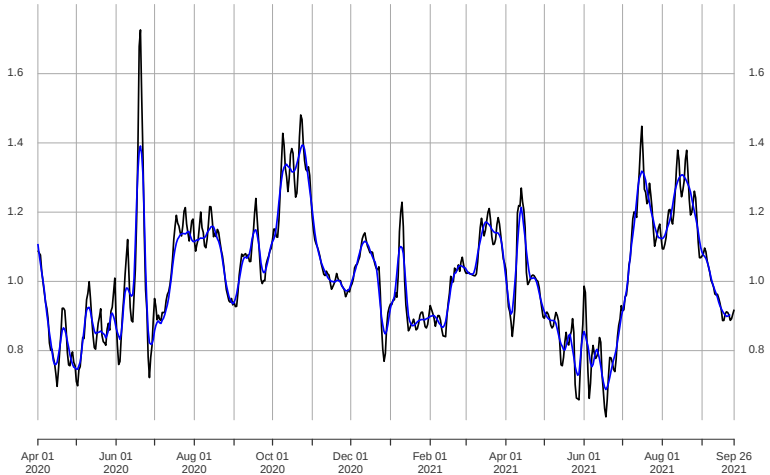


Figure: Schätzer für den Erwartungswert von R_{t-14} .

Modellanpassung aus zur Verfügung stehenden Daten

Zur **Anpassung** einer **Negativen Binomialverteilung** mussten wir tiefer in die zur Verfügung stehenden Daten einsteigen.

Wir benötigen einen **Schätzer für die Streubreite** der Verteilung der individuellen Ansteckungsanzahlen. Dazu sind aber **Messwiederholungen** notwendig!

Um in gewisser Weise **Messwiederholungen** zu erhalten, haben wir aus den **Landkreisen** in Deutschland veröffentlichte Anzahlen positiv getesteter Personen verwendet, um eine Schätzung für die Streubreite der individuellen Ansteckungsanzahlen zu erhalten.

Gerade für die Möglichkeit der Einbeziehung von etwaiger **Überdispersion** ist die **Kenntnis der Streuung** unumgänglich.

Modellanpassung aus zur Verfügung stehenden Daten

Die Streuung der gemeldeten Anzahl von individuellen Ansteckungen hängt relativ kompliziert vom *reporting* ab (es ist nicht nur ein Faktor wie beim Erwartungswert).

Grund: Bei jeder sekundär infizierten Person entscheidet ein weiterer Zufallsprozess, ob PCR-getestet (und damit gemeldet wird) oder nicht.

↪ Komplexe Erhöhung der Variabilität, die sich nicht herauskürzt!

Wir nehmen Werte für Reporting-Rate p_0 an, um Streuung schätzen und darauf aufbauend Negative Binomialverteilung zur Beschreibung der Anzahl individueller Sekundärinfektionen R anpassen zu können.

Wir haben für verschiedene Werte von p_0 die folgenden Schätzungen der Streubreite der Verteilung von R erhalten:

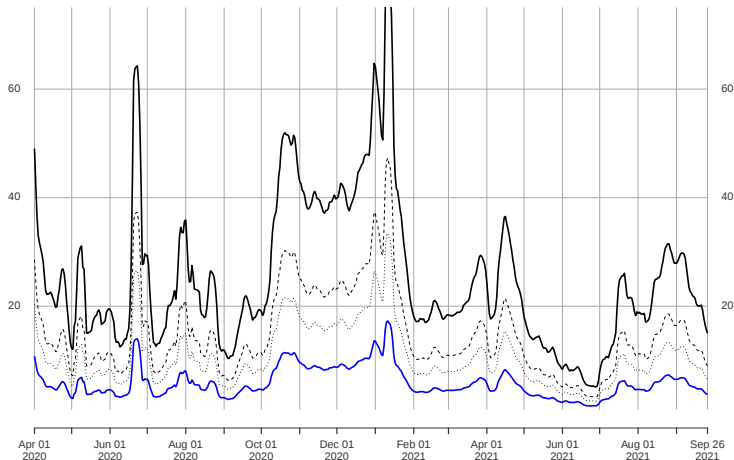


Figure: Schätzer für die Varianz von R_{t-14} für $p_0 = 0.2$ (durchgezogen), $p_0 = 0.35$ (gestrichelt), $p_0 = 0.5$ (gepunktet).

Zum Vergleich: Der 7-Tage R-Wert des RKI

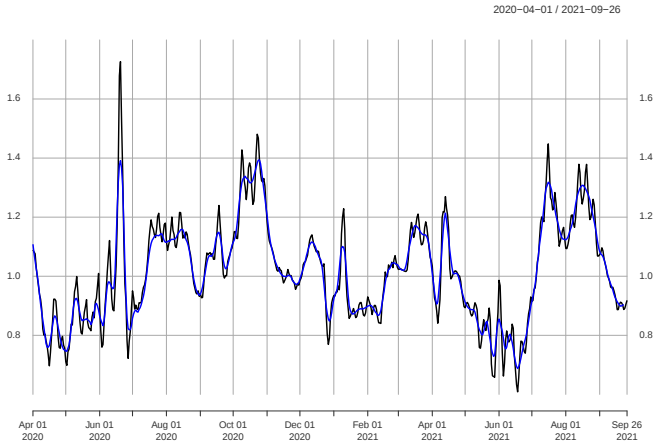


Figure: Schätzer für den Erwartungswert von R_{t-14} .

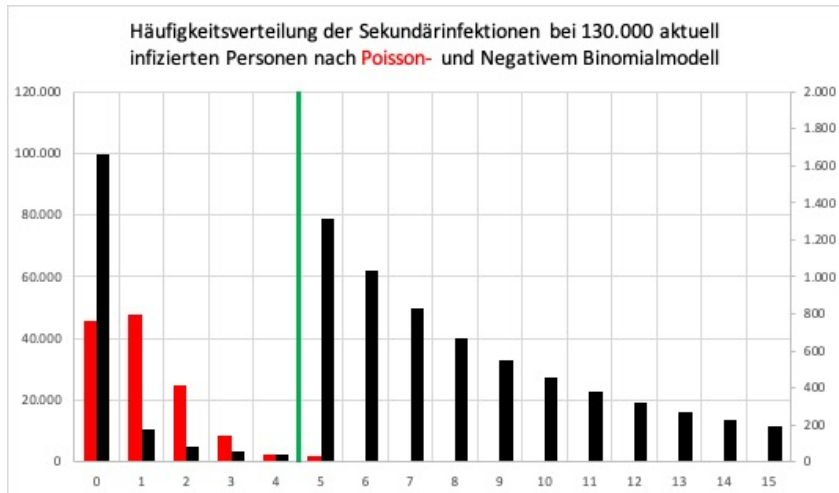
Die geschätzte Streuung der Verteilung von R

Gerade für die Hochphase der Infektion im Winter 2020/21 zeigt sich ein deutlich unterschiedliches Verhalten zwischen dem Erwartungswertschätzer (der 7-Tage R -Wert) und dem von uns aus Landkreisdaten entwickelten Streuungsschätzer.

Es sei erwähnt, dass bei Verwendung der **Poissonverteilung** Erwartungswert und Varianz vom Wert her identisch sein müssen, dass also kein unabhängiges Variieren voneinander darstellbar ist!

Übrigens: Die eingangs gezeigten Beispiele von Poisson- und Negativer Binomialverteilung sind typische konkrete Verteilungen, wie sie aus einer von uns durchgeführten Modellanpassung mit Hilfe der SARS-CoV-2-Infektionszahlen hervorgegangen sind. **Zur Erinnerung:**

Modellmäßige Häufigkeiten von Sekundärinfektionen



Modellrechnungen für Sekundärinfektionen

Mit Hilfe des angepassten Negativen Binomial-Modells für die individuellen Ansteckungszahlen können wir nun natürlich **Modellrechnungen** anstellen.

Etwa können wir leicht **modellmässige Wahrscheinlichkeiten** für bestimmte Anzahlen von individuellen Sekundärinfektionen berechnen. Diese Berechnungen liefern uns aber **keine Ursachen**, etwa für die Veränderungen der Wahrscheinlichkeiten über die Zeit!

Die folgenden Grafiken geben berechnete Wahrscheinlichkeiten für das Auftreten von folgenden Anzahlen von Sekundärinfektionen an:

- **keine** Sekundärinfektion
- **1-5** Sekundärinfektionen
- **mehr als 20** Sekundärinfektionen von einer einzigen primär infizierten Person

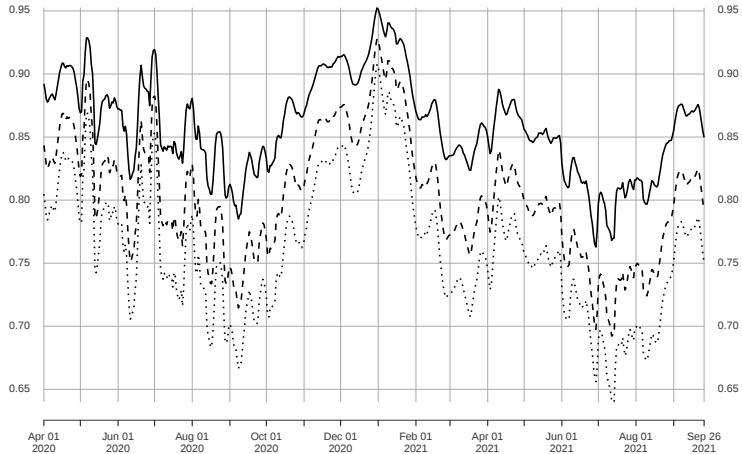


Figure: Wahrscheinlichkeit, dass ein zur Zeit $t - 14$ infizierter individueller Fall **keine** Sekundärinfektion hervorruft für $p_0 = 0.2$ (durchgezogen), $p_0 = 0.35$ (gestrichelt), $p_0 = 0.5$ (gepunktet).

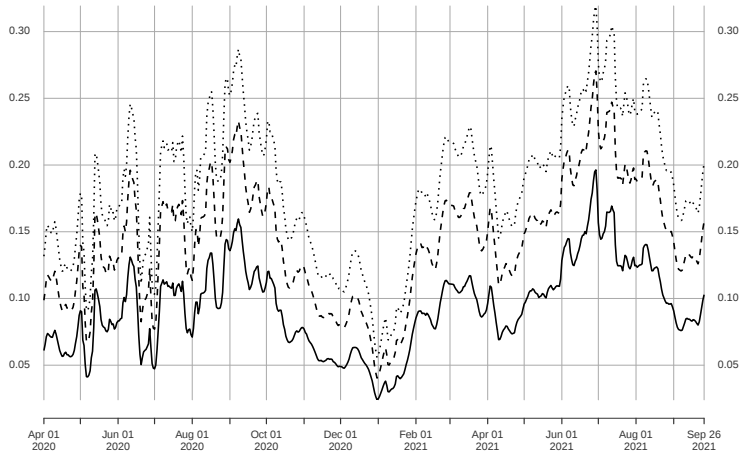


Figure: Wahrscheinlichkeit, dass ein zur Zeit $t - 14$ infizierter individueller Fall 1-5 Sekundärinfektionen hervorruft für $p_0 = 0.2$ (durchgezogen), $p_0 = 0.35$ (gestrichelt), $p_0 = 0.5$ (gepunktet).

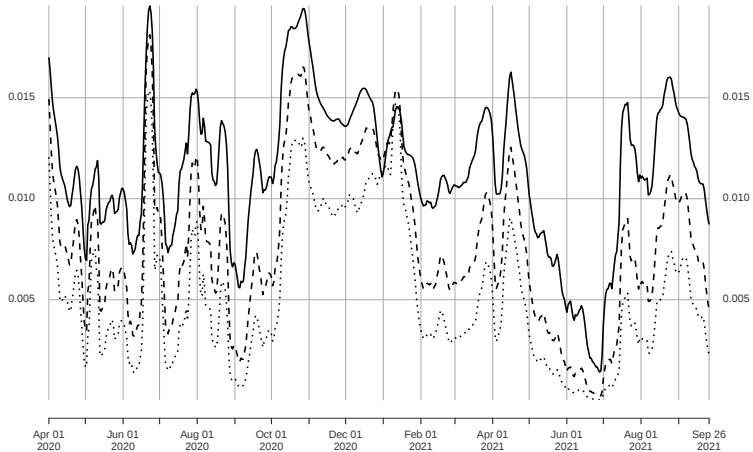


Figure: Wahrscheinlichkeit, dass ein zur Zeit $t - 14$ infizierter individueller Fall **mehr als 20** Sekundärinfektionen hervorruft für $p_0 = 0.2$ (durchgezogen), $p_0 = 0.35$ (gestrichelt), $p_0 = 0.5$ (gepunktet).

Zusammenfassung

Wir haben uns mit den folgenden Punkten auseinandergesetzt:

- Die individuelle Reproduktionszahl R als zufällige Größe
- Unterschiede verschiedener Wahrscheinlichkeitsmodelle für R
- Modellanpassung aus zur Verfügung stehenden Daten:
Der 7-Tage R-Wert des RKI und seine Bedeutung für eine Wahrscheinlichkeitsmodellierung von R
- Under-Reporting und Konsequenzen
- Das Streuverhalten der individuellen Reproduktionszahl und Überdispersion
- Modellrechnungen für das Auftreten von Sekundärinfektionen

Etwas Literatur

Braumann, A., Krampe, J., Kreiss, J.-P., Paparoditis, E. (2021). Estimation of the Distribution of the Individual Reproduction Number: The Case of the COVID-19 Pandemic.

Buitrago-Garcia et al. (2020). Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: *PLoS medicine* **17**(9).

Chowell, G. and Brauer, F. (2009). The Basic Reproduction Number of Infectious Diseases: Computation and Estimation Using Compartmental Epidemic Models. Springer, Dordrecht .

Cori, A. et al. (2013). A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *American Journal of Epidemiology* **178** .

Lloyd-Smith, J.O. et al. (2005). Superspreading and the Effect of Individual Variation on Disease Emergence. *Nature* **438**, 355–359.

Verlautbarung des Robert-Koch-Instituts (2020). Erläuterung der Schätzung der zeitlich variierenden Reproduktionszahl R. *Robert-Koch-Institut*.

Rahmandad, H. et al. (2020). Estimating COVID-19 Under-Reporting Across 86 Nations: Implications for Projections and Control. *MedRxiv* Preprint.

Santos-Hövenner, C. et al. (2020). Serology- and PCR-Based Cumulative Incidence of SARS-CoV-2 Infection in Adults in a Successfully Contained Early Hotspot (CoMoLo study), Germany, May to June 2020. *Euro Surveill.* **25**(47).

Streeck, H. et al (2020). Infection Fatality Rate of SARS-CoV-2 in a Super-Spreading Event in Germany. *Nature Communications* **11**.