

En matematisk model for hvordan en ny corona-variant bliver dominerende

Viggo Andreassen

Pandemix-centeret og IMFUFA

Department of Science. Roskilde University, DK-4000 Roskilde, Denmark

e-mail: viggo@ruc.dk

October 6, 2021

I løbet af 2021 har vi to gange oplevet, hvordan en ny variant af coronavirus er kommet til Danmark og i løbet af få måneder er blevet den dominerende virus variant. I denne tekst vil vi opstille en model, der viser, hvordan andelen af de smittede med den nye variant p udvikler sig, og fitte modellen til danske data.

Baggrund

I november 2020 opdagede de engelske epidemiologer, at der var opstået en ny variant – B.1.1.7 – af coronaviruset. Varianten opstod antagelig i Syd-øst England i september 2020, men blev først identificeret et par måneder senere. Det er den variant, som i dag kendes som **alfa-varianten**. Spredningen af den nye variant kunne følges, fordi ca. en tredjedel af de engelske PCR-prøver blev undersøgt for tilstedeværelsen af to forskellige corona-gener. Hos alfa-varianten var det ene af disse gener (S -genet) ændret så meget, at testen ikke længere var positiv for S -genet, mens prøverne stadig var positive for det andet gen. Epidemiologerne siger, at disse prøver udviste *S-gene Target Failure*, SGTF [Voltz et al., 2021]. Figuren viser den tidlige udvikling af andelen af prøver med SGTF af alle de prøver, som blev testet for begge gener.

I april 2021 dukker der atter en ny variant op. Det er B.1.617.2 kendt som **delta-varianten**. Delta-varianten har det oprindelige S -gen og udviser derfor ikke SGTF.

En model for variant overgangen

For at forstå hvordan hyppigheden af en ny variant udvikler sig, kan man tage udgangspunkt i den velkendte SIR -model, men med den tilføjelse, at sygdommen forekommer i to udgaver. Vi betegner antallet, som er smittet med den oprindelige udgave med I_W (Wuhan) mens antallet, der er smittet med den nye udgave betegnes I_B (B.1.1.7). Modellen er diskuteret i nogen detalje i et bilag til Pedersen et al. [2021].

Begge varianter af sygdommen kan smitte de modtagelige i befolkningen S , og er man først kommet sig efter infektion med en variant, er man immun R over for yderligere smitte fra begge varianter.

Med introduktionen af den nye variant bliver SIR -modellen til

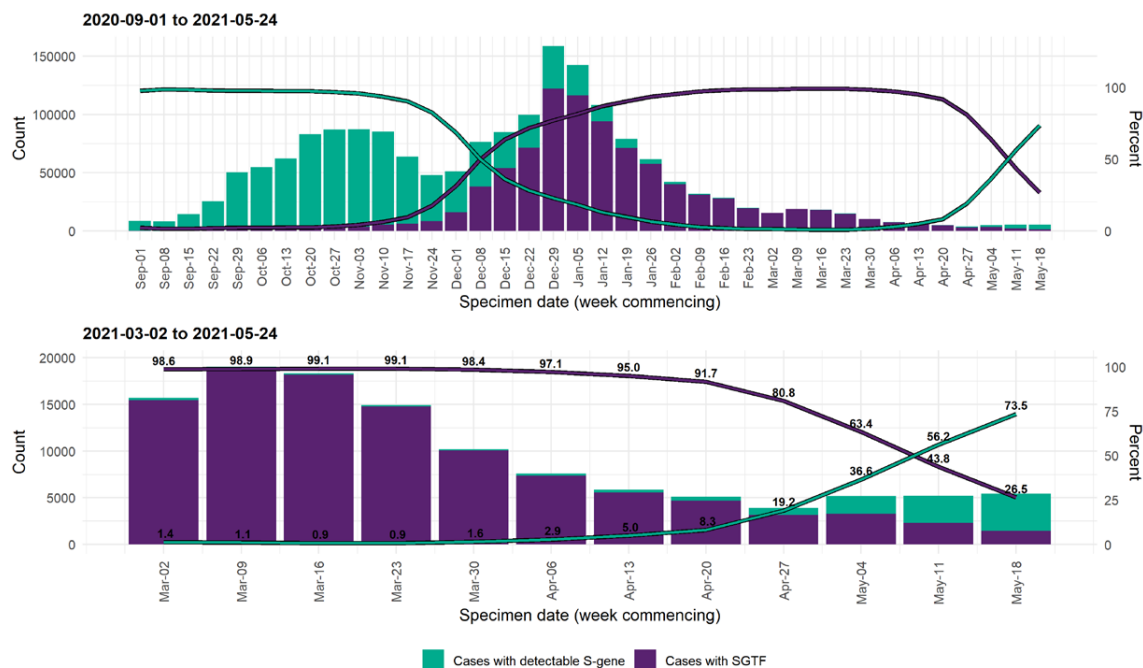
$$S'(t) = -\beta_W I_W(t)S(t) - \beta_B I_B(t)S(t) \quad (1)$$

$$I'_W(t) = \beta_W I_W(t)S(t) - \nu_W I_W(t) \quad (2)$$

$$I'_B(t) = \beta_B I_B(t)S(t) - \nu_B I_B(t) \quad (3)$$

$$R'(t) = \nu_W I_W(t) + \nu_B I_B(t) \quad (4)$$

Modellen kan kun løses ved simulering, men her vil koncentrere os om at forstå, hvordan hyppigheden af den nye variant udvikler sig, dvs vi fokuserer på $p = \frac{I_B}{I_W + I_B}$.



Only tests carried out with the TaqPath PCR assay and with confirmed SGTF or S gene positive results included, from Alderley Park, Milton Keynes and Glasgow Lighthouse Laboratories.
Case with SGTF: Positive SARS-CoV-2 test with non-detectable S gene and ≤ 30 CT values for N and ORF1ab genes.
Case with detectable S gene: Positive SARS-CoV-2 test with ≤ 30 CT values for S, N, and ORF1ab genes.
Data source: SGSS. Cases deduplicated to one positive test per person per week.

Figure 1: Overgangen fra den oprindelige Wuhan stamme til alfa-varianten og fra alfa-varianten til delta-varianten. Figuren viser andelen af positive prøver, der bliver testet i de engelske Lighthouse Laboratorier, og som som ikke indeholder **S**-genet. Efteråret 2020: Wuhan varianten har **S**-gen. Vinter 2020: B1.1.7 har ikke **S**-gen. Foråret 2021: B1.617.2 har **S**-gene. Fra [Public Health England, 2021, p 65]

Ved hjælp af regnereglerne for differentiation (af en brøk) og en pæn portion algebra kan man vise, at

$$p'(t) = ((\beta_B S(t) - \nu_B) - (\beta_W S(t) - \nu_w)) p(t)(1 - p(t)), \quad (5)$$

hvor $S(t)$ skal findes ved at løse ligningerne (1–4).

Det overlades til læseren at gå gennem detaljerne i udledningen af ligning (5), se opgave 1.

Simplifikation og analyse

I Danmark har corona-epidemien været undertrykt, sådan kun relativt få mennesker er blevet smittet. Undersøgelser af blodet fra bloddonorer viser, at i det sene forår 2021 havde kun ca 8% af befolkningen antistoffer over for corona. Det betyder, at vi kan tillade os at simplificere modellen ved at antage, at $S(t) = N$ er konstant gennem hele overgangsperioden.

Undersøgelser viser, at varigheden af smitte-forløbet er nogenlunde det samme for de tre varianter. Det betyder, at vi med god tilnærmelse kan antage at $\nu_W = \nu_B = \nu$.

Ved at indføre tærskelværdien

$$\mathfrak{R}_0 = \frac{\beta N}{\nu}$$

kan ligning (5) simplificeres til

$$p'(t) = \nu(\mathfrak{R}_{0B} - \mathfrak{R}_{0W})p(1 - p), \quad (6)$$

se opgave 2.

Bemærk, at ligningen nu kan løses uafhængigt af den fulde model.

Den matematiske analyse

Læseren kan nu gå to veje i analysen af model (6).

- Modellen kan simuleres f.eks. ved hjælp af Fpro
- Modellen kan løses ved hjælp af variabel separation. Her bevæger vi os ud over gymnasie-pensum idet man skal vide at

$$\int \frac{dp}{p(1-p)} = \ln(p/(1-p)).$$

Data fitnings-øvelsen i Opgave 3 kan løses ad begge veje.

Biologisk tolkning af modellen

(Afsnittet kan springes over ved første gennemlæsning.)

Ved et umiddelbart blik ligner ligning (6) modellen for logistisk vækst, hvis bærekapaciteten sættes til 1 (eller 100%). Fra et formelt matematisk synspunkt er dette korrekt, og det giver i en undervisningssammenhæng mulighed for at trække på evt erfaringer fra tidligere arbejde med logistisk vækst.

Set fra et biologisk synspunkt er det imidlertid ikke en frugtbar tolkning af (6). Modellen skal forstås i en *populationsgenetisk sammenhæng*. Og her er der desværre brug for yderligere et skridt for at vise sammenhængen.

Man skelner i populationsgenetik mellem modeller med diskrete generationer og modeller med overlappende generationer (kontinuerte modeller).

Modeller med diskrete generationer er langt de mest udbredte og de eneste, der egner sig til at beskrive situationer, hvor individet bærer to kopier af hvert gen (diploider). De fleste biologer kender kun denne såkaldte Wrights populationsgenetik. Fitness er defineret som *antallet af levedygtigt afkom per individ*

Virus har imidlertid kun en kopi af hvert gen (de er haploider). Her kan man som alternativ benytte den kontinuerte model (Malthus populationsgenetik) hvor fitness defineres som *vækstraten* for den pågældende genotype.

Modellen (6) kan nu opfattes som modellen for naturlig selektion mellem to genotyper som har (Malthus) fitness $\nu\mathfrak{R}_{0B}$ hhv $\nu\mathfrak{R}_{0W}$.

Tærskelværdien $\mathfrak{R}_0$ kan i øvrigt tolkes som *antallet af nye smittede, som en smittet giver anledning til når alle er modtagelige*. Det svarer således til Wrights fitness ved epidemi-start hvis virus optælles i enheder at smittede personer.

Danske data

I Danmark har vi ikke haft mulighed for at følge tabet af *S*-genet fordi vi anvender en anden analysemetode. Tilgængæld har vi i Danmark gen-sekventeret en meget stor del af de positive prøver. Oplysninger om de sekventerede prøver kan findes på hjemmesiden www.covid19genomics.dk, hvor man uge for uge kan se, hvor mange af de sekventerede prøver, der er af type B.1.1.7 (alfa) og – senere i forløbet – hvor mange prøver, der er af typen B.1.617.2 (delta). Tallene bliver løbende opdateret efterhånden som flere (gamle) prøver bliver færdig-sekventeret, så tallene i det medfølgende regneark vil ikke stemme helt overens med de tal man kan finde ved et senere down-load.

I regnearket **DKvariants.xlsx** kan man finde oplysninger om, hvor mange prøver der er blevet sekventeret fra hver uge, samt hvor mange af disse prøver, der er af type B.1.1.7 hhv B.1.617.2. I de fleste uger er ca 95% af alle positive prøver blevet type-bestemt. I de resterende prøver er genmaterialet for ufuldstændigt til at tillade en type-bestemmelse. Der er således tale om et meget mere komplet data-sæt end det engelske.

Opgave 1

Udled ligning (5) fra den store model (1–4). Vink: Start med at differentiere p ud fra definitionen dvs $p = \frac{I_B}{I_W + I_B}$. Indsæt de kendte udtryk for $I'_B(t)$ og $I'_W(t)$ fra den fulde model. Resten følger ved brug af almindelig regning.

Opgave 2

Udled ligning (6).

Opgave 3

Fit data fra de danske prøver for B.1.1.7 til model (6) og find $\mathfrak{R}_{0B}$, hvis vi antager at $\nu = 1/5.6$ dage og $\mathfrak{R}_{0W} = 2.4$.

Der er mindst 3 måder at gå til opgaven:

1. fit kurven vha Fpro;
2. fra variabel separation kan man se at

$$\ln(p/(1-p)) = a(t-t_0) + \ln(p_0/(1-p_0)),$$

hvor $a = \nu(\mathfrak{R}_{0B} - \mathfrak{R}_{0W})$. Man kan således klare sig med at fitte til en ret linje.

3. Plot $p(t)$ ud fra data, og find grafisk hældningen af kurven i punktet hvor $p = 1/2$.

Opgave 4

Gentag analysen for overgangen fra B.1.1.7 til B.1.617.2 i ugerne 2021-W18-32.

Der er en særlig udfordring her. Hyppigheden af corona varierer meget gennem perioden fordi sæson-effekten gør corona mindre smitsom, mens verdensmesterskaberne i fodbold trækker den anden vej, lige som samfundet lukkes mere op mod periodens afslutning. Endelig vokser andelen af vaccinerede betragteligt gennem perioden. Hvilken indflydelse har det på resultaterne?

Man kan finde en opgørelse over andelen af befolkningen, som har anti-stoffer over for Covid-19 på hjemmesiden bloddonor.dk/coronavirus for hver uge frem til 2021-W26.

References

- E. Voltz et al. Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. medRxiv, 2021.
- Pedersen, Andreasen, and Berrig. Britisk variant lurker stadig: Denne interaktive beregner viser, hvorfor vi er nødt til at holde fast. www.Videnskab.dk, 2021.
- Public Health England. SARS-CoV-2 variants of concern and variants under investigation in England, Technical Briefing 15, 11 June, 2021.