

The role of data science in encountering COVID-19 pandemic

Jamshidi A.A.¹ and Modaresi M.S.²

¹ School of Mathematics, Statistics and Computer Science, College of Science, University of Tehran

Advanced Systems Biology and Cancer Research Lab., Tehran, I.R. of Iran

² Biotechnology Department, College of Science, University of Tehran, Tehran, I.R. of Iran

Abstract

The COVID-19 pandemic started in China and quickly spread throughout the world. This pandemic has affected the health of millions of people, the global economy and has challenged the majority of scientists and politicians to find effective medical treatments and intelligent policies to reduce the side effects of this pandemic as soon as possible. Technological advancements have enabled us to collect tremendous amount of data in various related areas that includes; cellular level data (genomics, proteomics and transcriptomics) and epidemiological data (number of patients, the behaviors of people in critical situations). The analysis of these data sets could help us better understand and cure this disease and control its transmission. In this paper, we provide a comprehensive overview of various areas of data science and their role in encountering COVID-19. In addition, we process the new daily cases of this disease in frequency domain. Frequency domain helps us understand the effect of various important delays in this pandemic. Based on the frequency content of daily new cases of the COVID-19, in the countries that are affected the most, we could classify various countries into four major classes.

Key words: Covid-19, SARS-CoV2, Artificial intelligence, system biology, Time delay dynamic model.

یک مدل ریاضی برای همه گیری

علی کمالی نژاد*

تهران، دانشگاه تهران، پردیس علوم، دانشکده ریاضی، آمار و علوم کامپیوتر

چکیده

مدل سازی ریاضی به صورت فزاینده ای توسط پژوهشگران به عنوان یک ابزار تحقیقاتی مهم در کنترل بیماری های مسری مورد استفاده قرار می گیرد و در حال حاضر، مدل های ریاضی به طور گسترده ای برای ارزیابی اثربخشی راهبردهای کنترل بیماری کوید-۱۹ استفاده می شوند. هدف از این مقاله، مرور مختصر مدل همه گیری SIR است که یکی از مدل های بنیادی در نظریه ریاضی همه گیرشناسی است.

* نویسنده مسئول، پست الکترونیکی: kamalinejad.a@ut.ac.ir

مقدمه

همه گیرشناسی از طریق مطالعه مختصر مدلی در این حوزه و ضرورت استفاده از مدل های کمی در تحلیل نحوه شیوع بیماری کوید-۱۹^۱ است.

به علاوه مدلی که انتخاب شده است، با وجود آن که مدلی کلاسیک محسوب می شود اما با تقریبی می توان از آن برای تحلیل ریاضی برخی جنبه های شیوع بیماری کوید-۱۹ نیز استفاده کرد.

این نوشتار مقدمه ای مختصر و کوتاه بر نظریه ریاضی همه گیرشناسی^۱ است.

پیش از آغاز لازم است تا بر نکته ای اساسی، تاکید شود. مطالبی که در ادامه خواهند آمد از بخش های کلاسیک نظریه ریاضی همه گیرشناسی است و می توانید این موضوعات را در (Bjørnstad, 2018)، (Brauer, 2019)، (Li, 2018)، (Martcheva, 2015) و (Yan, 2019) بیابید. به این ترتیب، اصولاً چرا چنین متنی نگاشته شده است؟ دلیل، معرفی مقدماتی نظریه ریاضی

² COVID-19

¹ Epidemiology

زمان، مبتلا می‌شوند را با d نمایش می‌دهیم. لذا نرخ تغییرات رده (۱) برابر خواهد بود با:

$$S'(t) = -d$$

حال تلاش می‌کنیم تا ضابطه‌ای برای d بیابیم.

فرض می‌کنیم تعداد کل تماس‌های یک فرد بیمار در واحد زمان، متناسب است با کل جمعیت و لذا به ازای ثابتی مانند c برابر cN است. احتمال این که تماس حاصل شده با فردی از رده (۱) باشد، برابر است با $\frac{S}{N}$ لذا $cN \frac{S}{N}$ تعداد تماس‌های یک فرد بیمار با افراد سالم اما مستعد ابتلا به بیماری در طول یک واحد زمانی خواهد بود. البته معمولاً در بیماری‌های مسری، هر تماس بین یک فرد مستعد و یک فرد بیمار، لزوماً به بیماری فرد مستعد، منجر نمی‌شود. بیایید فرض کنیم که احتمال بیماری فرد سالم اما مستعد در اثر تماس با فردی بیمار p باشد. در این صورت $pcN \frac{S}{N} = pcS$ تعداد افراد سالم (اما مستعد به ابتلا به بیماری) است که در اثر تماس با یک فرد بیمار در یک واحد زمان بیمار می‌شوند. لذا $d = pcSI$ تعداد کل افرادی است که در یک واحد زمان بیمار می‌شوند.

در این صورت و از کل بحث فوق نتیجه می‌شود که:

$$S'(t) = -pcI(t)S(t)$$

و اگر از نماد $\beta = pc$ استفاده کنیم داریم:

$$S'(t) = -\beta IS$$

اکنون تلاش می‌کنیم تا ضابطه‌ای برای نرخ تغییرات رده (۲) یعنی I بیابیم. افراد سالم اما مستعدی که مبتلا می‌شوند از رده (۱) به رده (۲) منتقل می‌شوند. از طرفی گروهی که بهبود می‌یابند از رده (۲) خارج می‌شوند. فرض کنید این افراد با احتمال α در واحد زمان بهبود می‌یابند. به عبارت دیگر αI تعداد افراد بیماری است که در واحد زمان بهبود می‌یابند. در این صورت:

$$I'(t) = \beta IS - \alpha I$$

در پایان و برای محاسبه نرخ تغییرات رده (۳)، توجه کنید که افرادی که بهبود می‌یابند از رده مبتلایان خارج شده و به رده بهبود یافتگان منتقل می‌شوند. لذا:

$$R'(t) = \alpha I$$

بنابراین و با توجه به کل بحث فوق، مدل کرماک-مک‌کندریک به صورت سیستم معادلات دیفرانسیل عادی زیر قابل بیان است:

$$\begin{cases} S'(t) = -\beta IS \\ I'(t) = \beta IS - \alpha I \\ R'(t) = \alpha I \end{cases}$$

برای توصیف دینامیک شیوع بیماری، می‌توان پاسخ‌های دستگاه فوق را با در دست داشتن شرایط اولیه $S(0)$ ، $I(0)$ و $R(0)$

در ادامه، مدل کرماک-مک‌کندریک^۱ به عنوان منبع اولیه تمام مدل‌های SIR و مدل‌های مشابه، معرفی خواهد شد. مطالعه این مدل می‌تواند برخی از گام‌های اساسی در مدل‌سازی ریاضی همه‌گیری را روشن کند.

به دست آوردن مدل

بر اساس مدل کرماک-مک‌کندریک، زمانی که یک بیماری مسری در جمعیتی شیوع پیدا می‌کند، بیماری آن جمعیت را به سه رده بدون اشتراک، افزاز می‌کند:

(۱) رده افرادی که سالم‌اند، اما مستعد ابتلا به بیماری هستند. تعداد افراد این رده را با S نمایش می‌دهیم.

(۲) رده افرادی که بیمارند. در این مدل فرض می‌کنیم که این افراد ناقل بیماری نیز هستند که البته فرضی طبیعی است. تعداد افرادی که در این رده قرار می‌گیرند را با I نمایش می‌دهیم.

(۳) رده افرادی که بهبود یافته‌اند و دیگر به بیماری مبتلا نمی‌شوند. اندازه این رده را نیز با R نمایش می‌دهیم.

روشن است که اندازه این سه دسته در طول زمان و در جریان شیوع یک بیماری تغییر می‌کنند.

بنابراین S ، I و R توابعی برحسب زمان هستند. در نتیجه آن‌ها را به عنوان توابعی بر حسب زمان به ترتیب با $S(t)$ ، $I(t)$ و $R(t)$ نمایش می‌دهیم. حال فرض کنید که کل جمعیتی که شیوع بیماری در آن رخ داده است N باشد.

در این صورت:

$$N = S(t) + I(t) + R(t)$$

در ادامه دو فرض ساده را در نظر خواهیم گرفت:

نخست آن‌که ناقلین، مبتلایان هستند. دوم آن‌که اندازه کل جمعیت مورد مطالعه یعنی N در طول زمان تغییر نمی‌کند.

حال تلاش می‌کنیم تا به کمک معادلات دیفرانسیل با مشتقات معمولی، دینامیک هر رده را توصیف کنیم.

برای به دست آوردن معادلات دیفرانسیل، فرض‌های دیگری در مورد این که چگونه رده‌ها در طول زمان تغییر می‌کنند، در نظر خواهیم گرفت. این کار را با مطالعه رده (۱) آغاز می‌کنیم.

زمانی که یک فرد سالم ولی مستعد ابتلا به بیماری در تماس با فردی بیمار قرار گیرد، فرد سالم با احتمالی مبتلا می‌شود و از رده (۱) به رده (۲) منتقل می‌شود. تعداد افرادی که در یک واحد

¹ Kermack-McKendrick Model

$$S_{\infty} = \lim_{t \rightarrow \infty} S(t) \quad , \quad R_{\infty} = \lim_{t \rightarrow \infty} R(t)$$

از آنجا که $S(t)$ و $R(t)$ به ترتیب نزولی و صعودی هستند و به ترتیب از پایین و بالا کراندارند، لذا S_{∞} و R_{∞} همواره موجودند. S_{∞} تعداد افراد مستعدی را نشان می‌دهد که در پایان شیوع، هنوز بیمار نشدند. همان‌گونه که شواهد تجربی نشان می‌دهند، همه‌گیری‌ها لزوماً با بیمار شدن تمام افراد به پایان نمی‌رسند و در پایان، افرادی وجود دارند که در طول آن همه‌گیری هرگز بیمار نشدند.

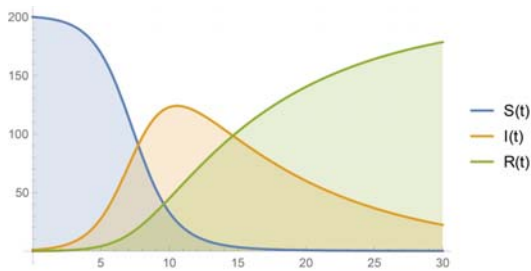
این مشاهده با مدل SIR نیز قابل تأیید است. برای محاسبه S_{∞} با توجه به معادلات مربوط به S و R در دستگاه SIR و قاعده مشتق‌گیری زنجیره‌ای داریم:

$$\frac{dS}{dR} = -\frac{\beta}{\alpha} S$$

با حل این معادله دیفرانسیل با شرط اولیه $S(0)$ داریم:

$$S = S(0)e^{-\frac{\beta}{\alpha}R} \geq S(0)e^{-\frac{\beta}{\alpha}N} > 0$$

در نتیجه $S_{\infty} > 0$



همچنین با توجه به مدل SIR همان‌گونه که انتظار داریم، در نهایت به پایان می‌رسد. برای مشاهده این موضوع، کافی است نشان دهیم که تعداد مبتلایان همواره به صفر خواهد رسید. برای اثبات، با انتگرال‌گیری در بازه $[0, \infty)$ از معادله

$$S'(t) = -\beta I(t)S(t)$$

داریم:

$$\int_0^{\infty} S'(t)dt = -\beta \int_0^{\infty} S(t)I(t)dt$$

$$S(0) - S_{\infty} = \beta \int_0^{\infty} S(t)I(t)dt$$

و از آنجا که $S(t)$ نزولی است، برای هر $t \in [0, \infty)$ و لذا: $S(t) \geq S_{\infty}$

$$S(0) - S_{\infty} \geq \beta S_{\infty} \int_0^{\infty} I(t)dt$$

در نتیجه $I(t)$ روی بازه $[0, \infty)$ انتگرال‌پذیر است و بنابراین:

$$\lim_{t \rightarrow \infty} I(t) = 0$$

مطالعه کرد. از این پس، مدل فوق را به اختصار مدل SIR می‌نامیم.

توجه کنید که واحد کمیت‌های سمت چپ معادلات فوق تعداد افراد در واحد زمان است. بنابراین طرف راست معادلات فوق نیز دارای همین واحد هستند. توجه به این موضوع کمک می‌کند تا واحدهای برخی ثابت‌های مدل را بیابیم. به عنوان مثال از معادله دوم مشخص می‌شود که واحد αN تعداد افراد در واحد زمان است و از آنجا که واحد N تعداد افراد است، لذا واحد α معکوس واحد زمان (یک تقسیم بر واحد زمان) است.

در ضمن با وجود این که افراد در طول زمان می‌تواند از رده‌ای به رده دیگر منتقل شوند، اما با توجه به یکی از مفروضات اولیه، انتظار داریم که تعداد کل جمعیت یعنی N ثابت بماند. اگر تعداد کل جمعیت در طول زمان را $N(t)$ بنامیم، آنگاه:

$$N(t) = S(t) + I(t) + R(t)$$

بنابراین با مشتق‌گیری نسبت به داریم:

$$N'(t) = S'(t) + I'(t) + R'(t)$$

از طرف دیگر و با توجه به دستگاه معادلات SIR اگر هر سه معادله را جمع کنیم، داریم:

$$S'(t) + I'(t) + R'(t) = -\beta IS + \beta IS - \alpha I + \alpha I = 0$$

در نتیجه $N(t)$ همواره مقدار ثابت

$$N(t) = N(0) = S(0) + I(0) + R(0) = N$$

است.

برخی خواص ریاضی SIR

در مدل SIR با توجه به این که همواره $S' < 0$ و $R' > 0$ لذا $S(t)$ و $R(t)$ به ترتیب نزولی و صعودی هستند. اما تعداد مبتلایان $I(t)$ می‌تواند دو وضعیت، داشته باشد: یا به صفر نزول کند و یا پس از صعودی به سطح ماکسیمم با نزولی به صفر برسد. تعداد مبتلایان در ابتدا صعود می‌کند، اگر

$$I'(0) = (\beta S(0) - \alpha)I(0) > 0$$

بنابراین شرط لازم و کافی برای افزایش اولیه در تعداد مبتلایان آن است که

$$\beta S(0) - \alpha > 0$$

یا به عبارت دیگر:

$$\frac{\beta S(0)}{\alpha} > 1$$

افزایش ناگهانی در تعداد مبتلایان و سپس کاهش آن‌ها روندی مشابه یک همه‌گیری است.

فرض کنید

یعنی نسبتی از افراد که تا زمان $t \geq 0$ هنوز مبتلا هستند برابر e^{-at} است. به تعبیر احتمالاتی، احتمال آنکه تا زمان t همچنان فردی بیمار باشد برابر e^{-at} است. به این ترتیب، نسبت افرادی که بهبود یافته‌اند برابر است با $1 - e^{-at}$. بنابراین تابع توزیع احتمال بهبودی عبارت است از:

$$F(t) = \begin{cases} 0 & t < 0 \\ 1 - e^{-at} & t \geq 0 \end{cases}$$

و تابع چگالی احتمال برابر است با:

$$f(t) = \begin{cases} 0 & t < 0 \\ ae^{-at} & t \geq 0 \end{cases}$$

در نتیجه متوسط زمان ماندن در رده مبتلایان برابر است با امید ریاضی زیر:

$$\int_{-\infty}^{\infty} tf(t)dt = \int_0^{\infty} tae^{-at}dt = \frac{1}{a}$$

به این ترتیب متوسط زمان بهبودی برابر است با $\frac{1}{a}$ و به کمک داده‌های مرتبط با زمان بهبودی بیماران، می‌توان برآوردی برای مقدار a به دست آورد.

برای برآورده کردن β و با توجه به معادلات دستگاه SIR ، مقدار a و مقادیر S_0 و I_0 می‌توان از روش حداقل مربعات برای یافتن بهترین برازش بر حسب β استفاده کرد و از این طریق برآوردی برای β بدست آورد.

SIR و شیوع بیماری کوید-۱۹

در ابتدا بر این نکته تأکید می‌شود که برای مطالعه نحوه شیوع بیماری کوید-۱۹ مدل‌های پیشرفته‌تر و به روزتری وجود دارند. اما مدل SIR پایه بسیاری از مدل‌های پیچیده‌تر امروزی است. به علاوه به نظر می‌رسد مفروضاتی که مدل را بر اساس آن‌ها به دست آورده‌ایم با تقریبی در مورد شیوع بیماری کوید-۱۹ نیز برقرارند. البته از آنجا که اطلاعات در مورد این بیماری، هنوز کامل نشده است، ممکن است که برخی از این مفروضات در آینده اعتبار خود را از دست بدهند.

فرض کنید قصد داشته باشیم تا از مدل SIR برای تحلیل دینامیک شیوع بیماری کوید-۱۹ در یک کشور استفاده کنیم. برای به دست آوردن مدل، فرض کردیم:

- تولد و مرگی در جمعیت رخ نمی‌دهد.

روشن است که این فرض، برقرار نیست. اما از آنجا که بازه مورد مطالعه این بیماری، حداقل در زمان نگارش این متن، چندان طولانی نیست، بنابراین تغییر جمعیتی که در اثر تولدها و مرگ‌های طبیعی ایجاد می‌شود، نسبت کوچکی از جمعیت کل است. البته مرگ‌هایی که در اثر ابتلا به این بیماری حادث می‌شوند، ممکن است قابل صرف‌نظر کردن نباشند. به این مورد

برای حل معادلات SIR ، نخست توجه کنید که R در دو معادله نخست دستگاه معادلات SIR ظاهر نمی‌شود و در صورت در دسترس بودن S و I می‌توان R را به کمک $R = N - S - I$ به دست آورد.

حال دستگاه معادلات دیفرانسیل زیر را در نظر بگیرید:

$$\begin{cases} S' = -\beta IS \\ I' = \beta IS - \alpha I \end{cases}$$

داریم:

$$\frac{I'}{S'} = \frac{\beta IS - \alpha I}{-\beta IS} = -1 + \frac{\alpha}{\beta S}$$

$$I' = \left(-1 + \frac{\alpha}{\beta S}\right) S'$$

و به کمک انتگرال‌گیری داریم:

$$I = -S + \frac{\alpha}{\beta} \ln S + C$$

بنابراین پاسخ ضمنی، با رابطه زیر به دست می‌آید:

$$I + S - \frac{\alpha}{\beta} \ln S = C$$

به عنوان کاربردی از پاسخ ضمنی فوق، می‌توان ماکسیمم تعداد افراد مبتلا را محاسبه کرد. در ماکسیمم I' ، یعنی:

$$\begin{aligned} 0 &= I' = \beta IS - \alpha I \\ S &= \frac{\alpha}{\beta} \end{aligned}$$

حال با توجه به شرایط اولیه $S_0 = S(0)$ و $I_0 = I(0)$:

$$I + S - \frac{\alpha}{\beta} \ln S = I_0 + S_0 - \frac{\alpha}{\beta} \ln S_0$$

و با جایگذاری $S = \frac{\alpha}{\beta}$ داریم:

$$I_{max} = -\frac{\alpha}{\beta} + \frac{\alpha}{\beta} \ln \frac{\alpha}{\beta} + S_0 + I_0 - \frac{\alpha}{\beta} \ln S_0$$

برآورد پارامترها

حال به این پاسخ این پرسش می‌پردازیم که پارامترهای مدل را چگونه می‌توان برآورد کرد؟

برای برآورد α فرض کنید در زمان صفر، I_0 مبتلا موجود باشد و فرض کنید در طول زمان، به تعداد مبتلایان فردی اضافه نمی‌شود. در این صورت دینامیک تعداد مبتلایان، با معادله دیفرانسیل زیر توصیف می‌شود:

$$I'(t) = -\alpha I, \quad I(0) = I_0$$

با حل این معادله داریم:

$$\begin{aligned} I(t) &= I_0 e^{-\alpha t} \\ \frac{I(t)}{I_0} &= e^{-\alpha t} \end{aligned}$$

رده‌ای مجزا برای فوت شدگان در نظر نگرفته‌ایم. اما توجه کنید که بهبود یافتگان، رده‌ای را در جمعیت مورد مطالعه تشکیل می‌دادند که دیگر قادر به مبتلا کردن فرد مستعدی نیستند و خود نیز دوباره به بیماری مبتلا نخواهند شد. این موضوع حتی در مورد فوت شدگان نیز صادق است. در حقیقت با به دست آوردن نرخ مرگ و میر بیماری کوید-۱۹ به کمک داده‌های آماری، می‌توان تعداد فوت شدگان بر اثر ابتلا به این بیماری را از طریق تعداد S ، I و R به دست آورد.

بنابراین با استفاده از داده‌های آماری و تخمین مقادیر α و β می‌توان با تقریبی از مدل SIR برای تحلیل شیوع بیماری کوید-۱۹ استفاده کرد. به خصوص به کمک نتایج به دست آمده از این مدل می‌توان مشاهده کرد که اگر هیچ اقدام موثری در جهت جلوگیری از شیوع بیماری انجام نگیرد یا روش درمانی موثری یافت نشود، باید خود را برای چه وضعیتی آماده کنیم و در آینده نه چندان دور چه روی خواهد داد؟ این خود به تنهایی، اطلاع ارزشمندی است که استفاده از مدل SIR را توجیه پذیر می‌کند. به این معنا که این مدل، حداقل یکی از الگوهای شیوع ممکن در آینده را پیش‌بینی می‌کند.

در ادامه خواهیم پرداخت. بنابراین اگر این دسته از مرگ‌ها را در نظر بگیریم فرض حادث نشدن تولد و مرگ در جمعیت مورد مطالعه با تقریبی برقرار است.

- ورودی و خروجی در جمعیت کل رخ نمی‌دهد.

این فرض هم به دلیل طولانی نبودن بازه زمانی مورد مطالعه و متوقف بودن و یا بسیار محدود بودن مهاجرت در زمان شروع بیماری کوید-۱۹، فرضی است که با تقریب خوبی برقرار است.

- بهبود یافتگان در مقابل بیماری مصونیت پیدا می‌کنند.

در این مورد هنوز نظری قطعی از طرف نهادهای پزشکی معتبر مانند سازمان جهانی بهداشت، ابراز نشده است. اما تعداد افرادی که برای بار دوم به این بیماری مبتلا شده‌اند، عدد بزرگی نیست. البته ممکن است که با گذشت زمان این فرض نادرست از کار دربیاید. اما با یافته‌های موجود، این فرض نیز با تقریبی برقرار است.

موردی که نیاز به اندکی توضیح دارد، دینامیک تعداد افرادی است که بر اثر ابتلا به این بیماری از دنیا می‌روند. در این مدل فرض کرده‌ایم که پس از مدتی یک فرد مبتلا، بهبود می‌یابد و

منابع

- Bjornstad, O. N. (2018). *Epidemics: Models and Data using R*. Springer.
- Brauer, F., Castillo-Chavez, C., & Feng, Z. (2019). *Mathematical Models in Epidemiology*. Springer.
- Li, M. Y. (2018). *An Introduction to Mathematical Modeling of Infectious Diseases*. Springer.

- Martcheva, M. (2015). *An Introduction to Mathematical Epidemiology*. Springer.
- Yan, P., Chowell, C. (2019). *Quantitative Methods for Investigating Infectious Disease Outbreaks*. Springer.



Hocosoconnect. 21/05/2020