コロナウイルス感染者のいる確率

新居 俊作 (九州大学数理学研究院)

2020 年 4 月 8 日

コロナウイルスの感染を避けるためイベントやパーティー等の人が集まる事は避けるようにとの要請が政府等からされていますが、実際それはどの程度危険なのかを計算してみます。具体的には、人の集まりに感染者がいる確率を集まりの人数によって表します：

- 東京都で隔離されていない感染者が千人いる場合、満員の山手線の一編成に感染者が乗っている確率は 50%前後。
- 大雑把に確率を考える時は比例的に考えて：

$$\text{感染者のいる確率} = \text{その場にいる人数} \times \frac{\text{感染者数}}{\text{全人口}} \tag{1}$$

  と思いがちですが[a] 実際にはこの様に計算した値は正しい値と比べて (場合によってはかなり) 小さいです。
- それにもかかわらず、特定の条件が満たされている時は、上式からそれなりにいい近似値が得られます。

[a]確率において比例的な考え方をすると間違えることは以下の様な簡単な例からも分かります。

表と裏が確率 $\frac{1}{2}$ で出るコインが二枚ある時、それらを投げて、表が一枚と裏が一枚出る確率は $\frac{1}{2}$ です。しかし、四枚を投げて、表と裏が二枚ずつ出る確率は $\frac{1}{2}$ ではありません。

以下では次の記号を使います：

N： 考えている地域の総人口

n： 感染者数 (隔離されていない保菌者数)

x： 集団の人数

また、次の記号は計算やグラフを書かせるアプリの入力に便利です (一応慣れていなくても本論は読めます。)：

$a_1, a_2, \ldots, a_n$ を数とする時、記号 $\displaystyle\prod_{k=1}^{n} a_k$ を次の右辺の意味で使う。

$$\prod_{k=1}^{n} a_k = a_1 \cdot a_2 \cdot \cdots \cdot a_n \qquad (\ a_1, a_2, \ldots a_n \text{の積}\ )$$

全人口が N 人でそのうち n 人がコロナに感染しているが未だ潜伏期間中で自覚症状がない等の理由で隔離されていないとします。

この状況でイベント、パーティー、仕事等で x 人が集まったとします。この時、この k 人の中に感染者が一人も居ない確率はどうなるでしょうか？

もし、x が N や n に比して非常に小さい場合は、感染していない人数は $N-n$ なので、一人々々が感染していない確率を単純に大体

$$\frac{N-n}{N} \tag{2}$$

として、x 人の中に感染者がいない確率を

$$\left(\frac{N-n}{N}\right)^x \tag{3}$$

としてよいでしょう。

しかし、x が大きくなるとそういう訳には行きません。それでもう少し丁寧に考えてみましょう。

集まった x 人を順番に並べます。先ず一人目が感染していない確率は

$$\frac{N-n}{N} \tag{4}$$

です。次に二人目が感染していない確率は、二人目が、感染していない $N-n$ 人のうち最初の一人以外の $N-n-1$ 人から選ばれていなければならないので

$$\frac{N-n-1}{N} \tag{5}$$

となります。同様に三人目が感染していない確率は、三人目が、感染していない $N-n$ 人のうち最初の二人以外の $N-n-2$ 人から選ばれていなければならないので

$$\frac{N-n-2}{N} \tag{6}$$

となります。同様に 3 人目、4 人目、. . .、x 人目まで考えると、k 人目が感染していない確率は

$$\frac{N-n-k+1}{N} \qquad (1 \leqq k \leqq x) \tag{7}$$

となります。

従って、x にの中に一人も感染者がいない確率は

$$\frac{N-n}{N}\cdots\frac{N-n-k+1}{N}\cdots\frac{N-n-x+1}{N}=\prod_{k=1}^{x}\frac{N-n-k+1}{N} \tag{8}$$

です。

従って、x 人の中に一人でも感染者がいる確率は、1 からこれを引けば良いので次の様になります：

$$y=1-\prod_{k=1}^{x}\frac{N-n-k+1}{N} \tag{9}$$

これをて計算して分析するのは難しいので、パソコンのアプリを使いましょう。Mac には Grapher というアプリがるのでそれを使ってグラフを書いてみました。

全国

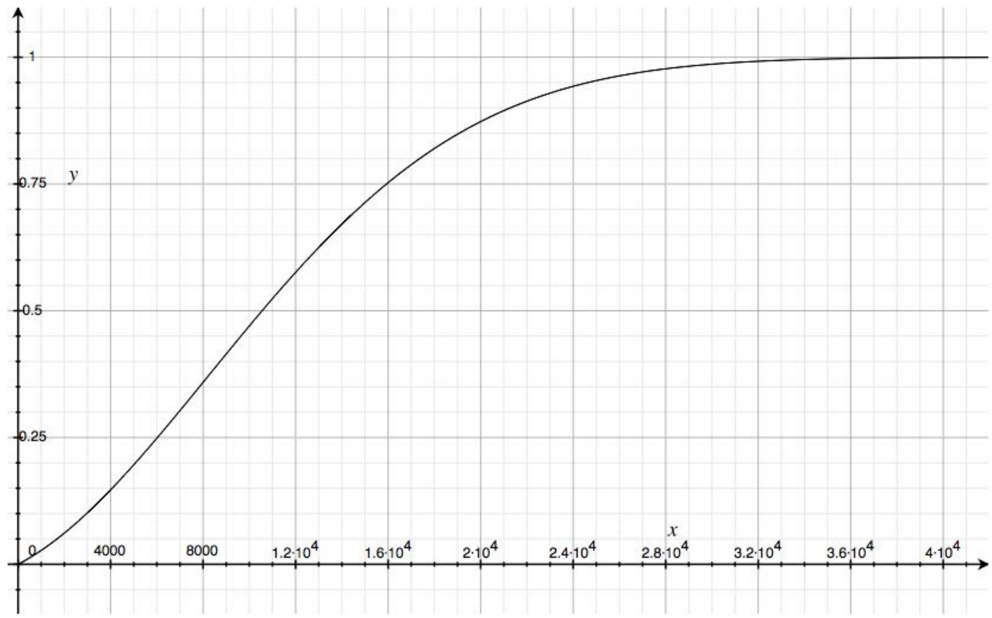
2020 年時点で日本の人口は 12600 万人居ます。累積感染者は 4 月 6 日時点で 4300 人居ますが、当然これは感染がわかった結果隔離されている人数なので、この人達が他にうつす可能性は低いはずです。なので潜伏期間が 2 週間程度あることを考えて、2 週間前の 3 月 25 日の感染者数 1300 人を引いた 3000 人が過去 2 週間、未だ発症していなかった為に感染していると分からずに周囲にうつしたと考えましょう。

従って、$N=12600000$ $n=3000$ で (9) は次のグラフの様になります。

$$y=1-\prod_{k=1}^{x}\frac{125997000-k+1}{126000000} \tag{10}$$

$x=40000$ 付近までのグラフを Graphar で書いたのが次です。

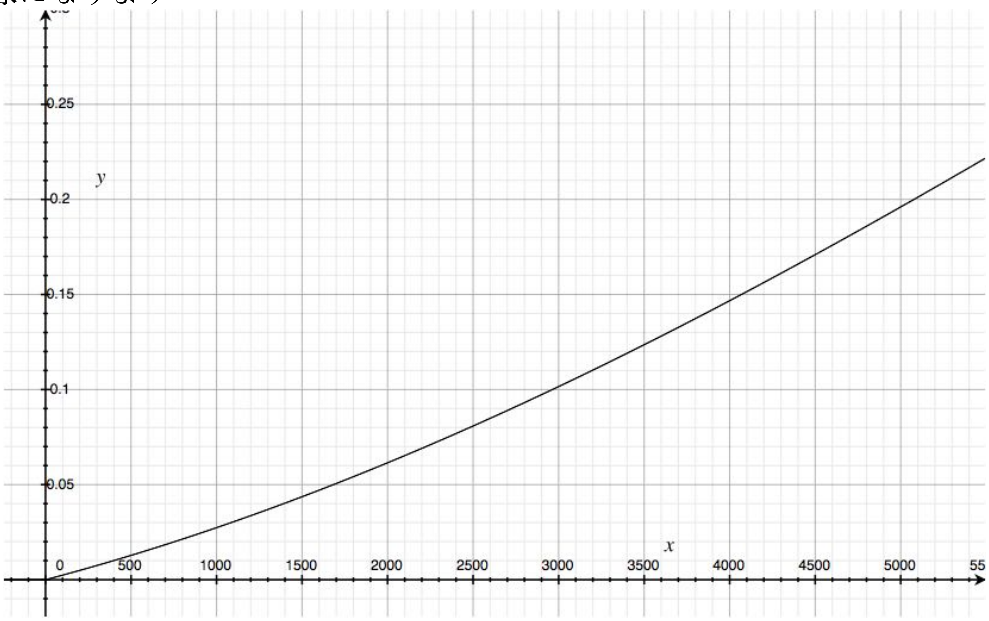
これを見ると、x が小さいうちはほぼ 0 です。また一人が感染している確率 $\dfrac{n}{N}$ は 4 万 2 千分の 1 なので、比例的に考えるとほぼ確実に感染者がいるのは 4 万人以上となりそうですが、x が 3 万位で確率はほぼ 1 になっていて、**比例的に考えるよりかなり小さな x でほぼ確実に感染者がいることになります。**



確率が４万２千分の１というと、「４万２千分人集めれば必ず一人、そして一人だけ」いる、と勘違いする人がいますが、そんなことはありません。その時によって、一人もいないこともあれば何人もいることもあります。「４万２千分人集めれば必ず一人、そして一人だけ」居る、のであれば、それはランダムに選んでおらず、作為的に選んでいるはずです。

「確率４万２千分の１」とは、一人ランダムに選んで感染しているかどうか確かめた後他の人と一緒にして、再びランダムに一人選んで確認する、ということを無限に行うと、平均で４万２千回に一回感染者が選ばれる、とうことです。

右下を拡大して $x = 5000$ くらいまでのグラフを書いてみると下のグラフの様になります：



$x = 5000$ を見ると、$y = 0.2$ 程度、すなわち 20% 程度で、単純に比例の感覚で考える 4 万 2 千分の 1 の 5000 倍、すなわち $\frac{5000}{42000} = 0.119$ よりかなり大きくなっています。

東京都

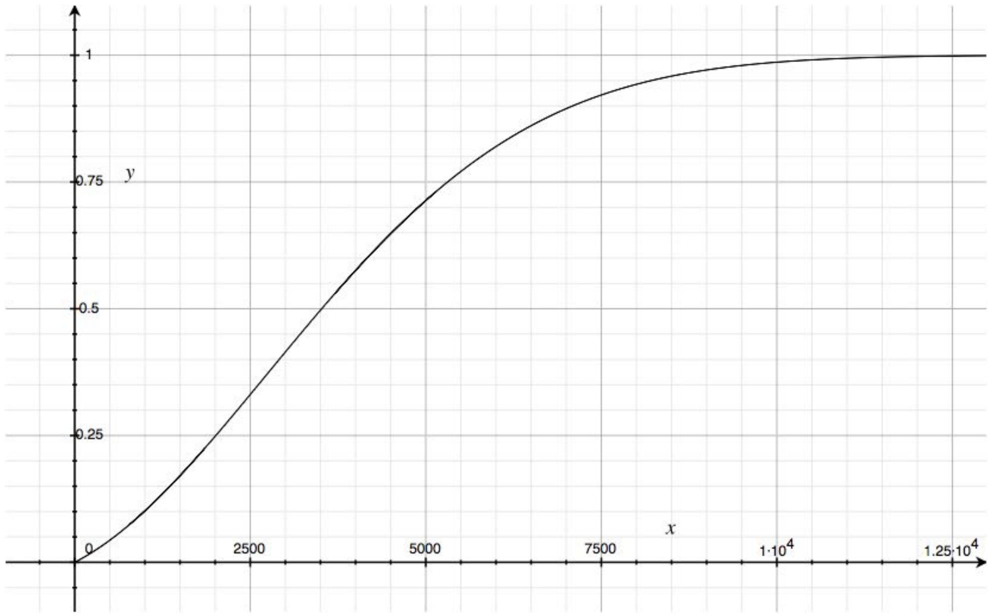
次に東京都について考えてみましょう。

2020 年時点で東京都の人口は 1400 万人居ます。累積感染者は 4 月 6 日時点で 1200 人で 2 週間前の 3 月 25 日の感染者数は 200 人です。なので、過去 2 週間に周りにうつしていた感染者数は 1000 人とし、$N = 14000000$ $n = 1000$ と考えましょう。

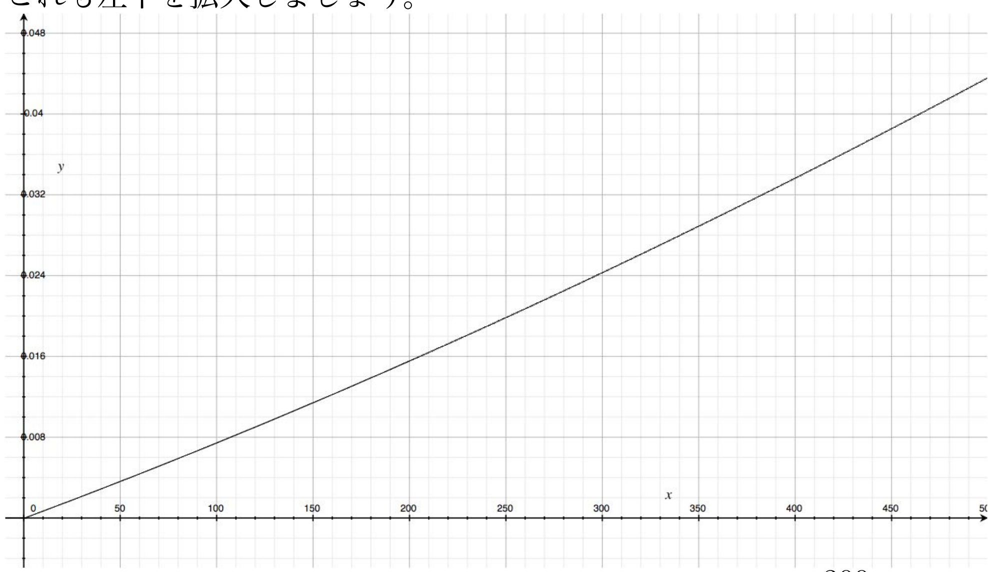
$$y = 1 - \prod_{k=1}^{x} \frac{13999000 - k + 1}{14000000} \tag{11}$$

$x = 13000$ 付近までのグラフを Graphar で書いたのが次のグラフです。

これを見ると、やはり x が小さいうちはほぼ 0 です。一人が感染している確率 $\frac{n}{N}$ が 1 万 4 千分の 1 ですが、そこまで行く前に x が 1 万で確率はほぼ 1 になります。なので**武道館で 1 万人集めると、ほぼ確実に感染者が居ます。**



これも左下を拡大しましょう。



$x = 300$ の時は、比例感覚で考える 1 万 4 千分の 1 の 300 倍は $\dfrac{300}{14000} = 0.021$ ですが、グラフを見ると $y = 0.024$ 程度、すなわち 2.4% 程度でやはり少し大きくなっています。満員電車の一両に乗っているのが 300〜350 人なので (国土交通省によると、1 両に 300 人で「体が触れ合い、相当な圧迫感がある。しかし、週刊誌ならなんとか読める」程度)、**東京の通勤電車で同じ車両に感染者がいる確率は 2.5〜3% 程度と考えられます。朝晩二回、二週間で 20**

回電車に乗ると思うと、決して小さくないと思われます。また、11 両編成の山手線一編成には 3500 人前後が乗っているので、グラフを見ると 50%前後の確率、言い換えれば電車 2 本に 1 本の確率で、同じ電車に感染者が乗っています。(これは比例的に考えるよりかなり大きいです。) もっとも、たとえ同じ電車に乗っていても、車両が違えば感染する危険は無いかもしれないので、これはあまり気にする必要はないかもしれません。

注意

全国の場合も東京都の場合も、**n が N に比べて非常に小さく、x も n および $\frac{N}{n}$ と比べて非常に小さい場合**は、比例の感覚

$$y \fallingdotseq x\frac{n}{N} \tag{12}$$

は、当たらずと言えども遠からずという感じになってます。これは以下の様に説明できます。

先ず (7) は次の不等式を満たします：

$$\frac{N-n}{N} \geqq \frac{N-n-k+1}{N} \geqq \frac{N-n-x}{N} \tag{13}$$

従って (8) はつぎを満たします：

$$\left(\frac{N-n}{N}\right)^x > \prod_{k=1}^{x} \frac{N-n-k+1}{N} > \left(\frac{N-n-x}{N}\right)^x \tag{14}$$

よって求める確率 (9) については

$$1-\left(\frac{N-n}{N}\right)^x < y = 1-\prod_{k=1}^{x} \frac{N-n-k+1}{N} < 1-\left(\frac{N-n-x}{N}\right)^x \tag{15}$$

となります。

高校で習う二項定理で最左辺を展開すると：

$$\begin{aligned}
1-\left(\frac{N-n}{N}\right)^x &= 1-\left(1-\frac{n}{N}\right)^x \\
&= 1-\left\{1-{}_xC_1\left(\frac{n}{N}\right)+{}_xC_2\left(\frac{n}{N}\right)^2+\cdots\right. \\
&\qquad \left.\cdots+(-1)^{x-1}{}_xC_{x-1}\left(\frac{n}{N}\right)^{x-1}+(-1)^x\left(\frac{n}{N}\right)^x\right\} \qquad (16) \\
&= x\left(\frac{n}{N}\right)-\frac{x(x-1)}{2}\left(\frac{n}{N}\right)^2+\cdots+(-1)^{x+1}\left(\frac{n}{N}\right)^x
\end{aligned}$$

同様に最右辺を展開すると：

$$\begin{aligned}1-\left(\frac{N-n-x}{N}\right)^x &= 1-\left(1-\frac{n+x}{N}\right)^x \\ &= 1-\left\{1-{}_xC_1\left(\frac{n+x}{N}\right)+{}_xC_2\left(\frac{n+x}{N}\right)^2+\right. \\ &\qquad \cdots+(-1)^{x-1}{}_xC_{x-1}\left(\frac{n+x}{N}\right)^{x-1} \\ &\qquad\qquad \left.+(-1)^x\left(\frac{n+x}{N}\right)^x\right\} \\ &= x\left(\frac{n+x}{N}\right)-\frac{x(x-1)}{2}\left(\frac{n+x}{N}\right)^2+\cdots \\ &\quad \cdots+(-1)^{x+1}\left(\frac{n+x}{N}\right)^x\end{aligned} \tag{17}$$

よって、先ず x が n と比べて非常に小さいければ、

$$\frac{n}{N} \fallingdotseq \frac{n+x}{N} \tag{18}$$

となって最右辺と最左辺がほぼ等しいので、結局

$$y \fallingdotseq x\left(\frac{n}{N}\right)-\frac{x(x-1)}{2}\left(\frac{n}{N}\right)^2+\cdots+(-1)^{x+1}\left(\frac{n}{N}\right)^x \tag{19}$$

となります。

更に、n が N に比べて非常に小さいと $\frac{n}{N}$ は非常に小さく、その結果 $k=2,3,\ldots,x$ について $\left(\frac{n}{N}\right)^k$ は $\frac{n}{N}$ に比べて非常に小さくなります。例えば $\frac{n}{N}=\frac{1}{1000}$ なら $\left(\frac{n}{N}\right)^2=\frac{1}{1000000}$ です。

そして x が $\frac{N}{n}$ に比べて非常に小さいならば

$$\begin{aligned}{}_xC_k\left(\frac{n}{N}\right)^k &= \frac{x(x-1)\cdots(x-k+1)}{(x-k)(x-k-1)\cdots2\cdot1}\cdot\frac{n}{N} \\ &< \frac{x^k}{(x-k)(x-k-1)\cdots2\cdot1}\left(\frac{n}{N}\right)^k \\ &< x^k\left(\frac{n}{N}\right)^k=\left(\frac{x}{(N/n)}\right)^k\end{aligned} \tag{20}$$

は $k=2,3,\ldots,x$ については $x\dfrac{n}{N}=\dfrac{x}{(N/n)}$ と比べて非常に小さくなります。従って、これらの和も小さいので、結果として (19) より

$$y \fallingdotseq x\left(\frac{n}{N}\right) \tag{21}$$

となり、この場合は (12) は正当化されます。