

スーパー・スプレッダーに関する考察。

コロナなど感染症の流行について、スーパー・スプレッダーの存在が指摘されている。感染症については、接触感染、空気感染、飛沫感染など感染のメカニズムが研究されているが、感染した者のすべてが、濃厚接触者のすべてに感染させるわけではない。感染される側については、感染しやすさが一様ではなく、感染しやすい人と、感染しにくい人が居るらしいことは、経験的な事実として納得できる。感染させる側についても、感染させる力が人によって違うらしい。感染させる力が強い感染者を、スーパー・スプレッダーと呼ぶ。

集団の中で、一人の感染者が何員の人を感染させるか、その人数を有効再生産数（R）という。有効再生産数は一種の期待値だから、実際に何員の人を感染させるかその確率は、集団の中で感染が起こる確率はそんなに高くはないだろうから、実際に何人を感染させるかその確率はポアソン分布的かもしれない。ポアソン分布では、期待値 λ の現象が k 回起こる確率は以下の式で与えられる。

$$P(k / \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$$

たとえば、有効再生産数3の時に、一人の人が k 人感染させる確率は下の表のように計算できる。

$\lambda = 3$	k	0	1	2	3	4	5	6	7	8	9	10
	λ^k	1	3	9	27	81	243	729	2187	6561	19683	59049
	k!	1	1	2	6	24	120	720	5040	40320	362880	3628800
	$e^{-\lambda}$	0.049787	0.049787	0.049787	0.049787	0.049787	0.049787	0.049787	0.049787	0.049787	0.049787	0.049787
		0.049787	0.149361	0.224042	0.224042	0.168031	0.100819	0.050409	0.021604	0.008102	0.002701	0.00081
	累積	0.049787	0.199148	0.42319	0.647232	0.815263	0.916082	0.966491	0.988095	0.996197	0.998898	0.999708
	1-累積	0.950213	0.800852	0.57681	0.352768	0.184737	0.083918	0.033509	0.011905	0.003803	0.001102	0.000292

つまり、65%の人は、3人以下しか感染させないが、35%の人は4人以上の人を感染させない。6人以上感染させる人は、3%にすぎない。つまり、30人に一人しかいない。私は、そのような人をスーパー・スプレッダーと呼ぶことに依存ないが、極論かもしれない。1%の確率でしか起こらないほどの人数の人に感染させる感染者をスーパー・スプレッダーと定義すると8人以上を感染させた人がスーパー・スプレッダーとなる。

このような考え方で、スーパー・スプレッダーを統計数学的に定義しようという提案は実際にある。

ここでは、感染者が濃厚接触する人の人数がわからないので、有効再生産数Rを期待値 λ として、ポアソン分布を考えて、スーパー・スプレッダーを定義した。人に感染させるのは感染後4日間だと考えて、4日間の濃厚接触者の数を n とする。たとえば職場の人数とか、4日間その店に来た人の総数を濃厚接触者の数 n と考えると良い。 n が大きくなれば、正規分布を仮定してもよさそうだが、 n が小さい場合もあると思うので、二項分布的に考える。

$$B(k/n) = \binom{n}{k} p^k (1-p)^{n-k}$$

濃厚接触者数 $n = 100$ として、感染する人の数の確率を二項分布的に計算してみた。

感染者数 k	0	1	2	3	4	5	6	7	8	9	10	
二項係数 (100 k)	1	100	4950	161700	3921225	75287520	1.19E+09	1.6E+10	1.86E+11	1.9E+12	1.73E+13	
p^k	0.03	1	0.03	0.0009	0.000027	8.1E-07	2.43E-08	7.29E-10	2.19E-11	6.56E-13	1.97E-14	5.9E-16
$(1-p)^{n-k}$	0.97	0.047553	0.049023	0.050539	0.052102	0.053714	0.055375	0.057088	0.058853	0.060674	0.06255	0.064485
確率		0.047553	0.14707	0.225153	0.227474	0.170606	0.101308	0.04961	0.020604	0.007408	0.002342	0.000659
累積		0.047553	0.194622	0.419775	0.647249	0.817855	0.919163	0.968772	0.989376	0.996784	0.999126	0.999785
1-累積		0.952447	0.805378	0.580225	0.352751	0.182145	0.080837	0.031228	0.010624	0.003216	0.000874	0.000215

感染確率 $p = 0.03$ 、濃厚感染者数 100 とすると、やはり 8 人以上感染する率は 1% 以下である。

スーパー・スプレッダーとはどんな人かという議論がある。より多くの人と濃厚接触すれば感染させる人の数も多くなるし、より濃厚な接触をすれば感染させる人の割合も増す。相違行動の違いではなくて、そもそも、その人の感染させる力が強いのだという可能性もある。たとえば、何かの原因で、その人の中でコピーされるウィルスの数が多いのかもしれない。基礎疾患がある場合、ウィルスのコピー数が多くなるという報告はある。こういう場合、実験生物学者としては、濃厚接触者の数をそろえて実験をしたくなる。もちろん、人を使ってそんな実験はできないだろう。倫理的にも問題だし、間違いなく犯罪だ。しかし、動物実験ならば、何か考えられそうだ。

たとえば、人為的に感染させたマウス 1 個体を 10 個体のマウスと濃厚接触させ、感染したマウスの数を調べます。これを感染させたマウス 1 個体ごとに 10 回繰り返して、たとえば、以下のようなデータが出とする。

1 回目	感染個体数 0
2 回目	感染個体数 1
3 回目	感染個体数 2
4 回目	感染個体数 2
5 回目	感染個体数 1
6 回目	感染個体数 1
7 回目	感染個体数 0
8 回目	感染個体数 0
9 回目	感染個体数 3
10 回目	感染個体数 1

要約すると、1 個体も感染しないのが 3 回、1 個体感染するのが 4 回、2 個体感染するのが 2 回、3 個体感染するのが 1 回で、合計 11 個体が感染した。全部で 100 個体を濃厚接触させてるので、感染率は $11/100=0.11$ ぐらいだろうという予測がつくが、念のために最頻値を見ると 1 個体で、やはり感染率は 0.1 程度だろうと考えられる。念のために、10 回の試行で、一人も感染しないのが 3 回あった。つまり、一人も感染しない確率は 30%、

確率 0.30 の事象だ、

$$B(k/n) = \binom{n}{k} p^k (1-p)^{n-k}$$

の式に当てはめてみる、

$$0.3 = \binom{10}{0} p^0 (1-p)^{10} = (1-p)^{10}$$

対数をとって

$$\begin{aligned}\log_e 0.3 &= 10 \log_e (1-p) \\ -1.20397280432594 &= 10 \log_e (1-p) \\ -0.120397280432594 &= \log_e (1-p)\end{aligned}$$

指数関数をとって元に戻す

$$\begin{aligned}e^{-0.1204} &= 1-p \\ 0.8866 &= 1-p \\ p &= 0.1134\end{aligned}$$

となる。

一人に感染させ割合は 0.40 だったのだから

$$\begin{aligned}0.4 &= \binom{10}{1} p^1 (1-p)^9 = 10p^1 (1-p)^9 \\ 0.04 &= p^1 (1-p)^9\end{aligned}$$

これを試行錯誤的に解くと、 $p = 0.097$

2人に感染させる割合は 0.20 だったから

$$\begin{aligned}0.2 &= \binom{10}{2} p^2 (1-p)^8 = 45p^2 (1-p)^8 \\ 0.0044 &= p^2 (1-p)^8\end{aligned}$$

これを試行錯誤的に解くと、 $p = 0.108$

3人に感染させる割合は 0.10 だったから

$$\begin{aligned}0.1 &= \binom{10}{3} p^3 (1-p)^7 = 120p^3 (1-p)^7 \\ 0.000833 &= p^3 (1-p)^7\end{aligned}$$

これを試行錯誤的に解くと、 $p = 0.128$

となって、これらのデータから1回の濃厚接触によって感染させる確率は、平均的に見て0.11前後だと考えて間違いなからう。このようなことから、感染実験では、この接触試験では、感染者の数は二に項分布的で、感染者のスプレッダーとしての機能は、10回の感染実験の間、ほぼ一定なのだということがわかった。すこで、このような実験で出てくる、感染者が感染後の一定期間、感染者が感染させる機能は、ほぼ一定だと考えることに問題はない

と結論する。つまり、この実験で、得た、濃厚接触さ 100 人当たりの感染者数を、感染させる強さの指標として、感染率とすることに問題はなさそうだ。このように、感染力がわかったマウスについて、ウィルスのコピー数を比較したり、マウスにストレスを与えて、ウィルスのコピー数の異なるマウスを作りだして、様々な条件で、感染実験を行うことには、大きな意義があるだろうと思う。

調査法と検定法を考える

マングローブ林は、熱帯、亜熱帯の沿岸に発達した林で、干潮時には干出するが、満潮時には、根本まで海水が上がってくる。強い日射と複雑な構造による多様性と変動により、生産性が高く、生物多様性が高く、魚や無脊椎動物など複雑な生態系が発達する。私たちの研究室は、タイのトランのシカオというところで、マングローブ林を流れる川の生態調査をしていた。東南アジアのマングローブ生態系で、大きな社会問題の一つが、エビ養殖場の生態系に与える影響であった。一時期、東南アジアでは、沿岸帯のマングローブ林を破壊して、エビ養殖場を建設することが盛んにおこなわれていた。それ自体が環境破壊であるが、養殖場から排出される残餌や糞などが、周辺の生態系に与える影響が懸念されていた。生態系そのものは複雑で、特定の生物の生息数でその影響を論ずることは難しい。シカオのマングローブ林の海岸を歩いているときに、私たちは、あることに気が付いた。干潮時、マングローブ海岸の砂浜には、多くのコメツキガニがいて、砂の団子を作っていた。それは、マングローブ林の中の川の上流から下流まで、ほぼ一様に無数と言っていいほどの数で存在した。コメツキガニが、砂の団子を作るのは、砂粒の表面の有機物を食べるためだ。養殖場がなければ、コメツキガニは、砂の上で増殖する付着珪藻などを食べているだろう。養殖場の周辺のコメツキガニは、養殖場から排出されるかなり細かい粒子として存在する残餌や糞を食べているはずだ。その当時、簡単に測定できる質量分析機が開発されて、様々な安定同位体比を使った、生態学的研究が流行り始めていた。養殖場で使われている配合飼料の原材料は、カタクチイワシの魚粉だ。付着珪藻は 1 次生産者で、カタクチイワシは植物プランクトン・動物プランクトンを食べている、2 次生産者以降の高次生産者だ。生体濃縮の結果、含まれている有機物の窒素や炭素などの安定同位体比が違うはずだ（わずかではあるが重さ違うので、代謝速度が違い、重いものが食性の段階ごとに濃縮されていく。）。当然、それらを食べているコメツキガニの体の安定同位体比は、養殖場のあるところと、ないところで違っているだろう。養殖場の周辺のコメツキガニの安定同位体比を測定すれば、養殖場が周辺のマングローブ生態系に与える影響を定量的に評価できるだろうというのが、私たちの最初の思い付きであった。それをどのように実現するかを考える。

もっとも単純に考えられるのは、養殖場の建設稼働の前後で、周辺のコメツキガニの安定同位体比を比較するという方法だ。この場合、養殖場 A の周辺に、サブ・サンプリングステーション a_1, \dots, a_n をもうけて、そのサブ・サンプリングステーションで 1 個体ずつ、コメツキガニを採集して分析し、前後で比較するというやりかただ。これだと、t 値としては

$$t = \frac{a_{k,after} - a_{k,before}}{\frac{s}{\sqrt{n}}}$$

という形になる、 s^2 としては、データから予測した普遍分散を使えばよいだろう。

$$s^2 = \frac{SS}{n-1}$$

$$SS = \sum_{k=1}^n (a_{k,after} - a_{k,before})^2 - \frac{1}{n} \left(\sum_{k=1}^n (a_{k,after} - a_{k,before}) \right)^2$$

これは、対になったデータ（前後で1対になっている）のt検定と呼ばれている検定の、もっともシンプルな形ですが、個体差を考えると、1サブステーションで1個体というのは、ちょっと危なそうだし、検出感度も悪そうだ。そこで、1サブステーションのサンプル数を3個以上にして平均化すれば、個体差が相殺されて、検出感度が上がる。

しかし、時間的な違いで、対をつくるというのは、危険かもしれない。シカオ周辺は、そのころ観光開発が盛んで、リボン島が過剰開発されて面白くなくなってきたので、プーケットなどの観光客が、結構、シカオ沿岸に表れるようになってきていた。干満差があるから、川の上流でも、その影響を受けている。すると、コントロール用のステーションとして、似たような環境のところに、ステーションBを設けて、そちらでも、対になったt検定をして、Aの変化が経年的な変動によるものかどうかを確かめた方が良いかもしれない（これは実験生態学ではよくある手法）。余裕があれば、全体の平均の値を期待値として、前後、ステーションという、2 X 2のカイ二乗検定をした方が丁寧かもしれないが、多分、論文としてはそこまで要求されない（論文の査読者ではないからはっきり言えない。簡単にできるから、要求があったら答えられるように、一様やって確認しておくということはありそう。）

現地の環境調査機関ならば、そういう機会があるかもしれないが、外国人の研究者には、養殖場の建設前後で調査できるという機会はあまりないだろう。短時間でやるには、時間的な変動ではなくて、空間的な変動を調べるしかない。そこで、考えられるのは、養殖場の影響が小さいと考えられるステーションをもうけて、そこと、ステーションAを比較することだ。具体的には、本流の流れによって、養殖場の影響が希釈されると考えられるステーションDをもうけて、サブ・サンプリングステーション d_1, \dots, d_n で採取を行って、Aとの違いを検討することを考えた。サブ・サンプリングステーションの採集個体数も複数個にした方が良いが、この場合、対になった、t検定は行えない。するとt値は

$$t = \frac{A_{mean} - D_{mean}}{\frac{s}{\sqrt{n}}}$$

とするしかないが、その場合、sをどうするか、nをどうするかという問題が生ずる。それを考えたのが、対にならないデータのt検定である。t検定の説明のほとんどはこのs、nをど

のように与えるかという話なのだが、結論は重み付き平均である。その詳細は、t 検定の講義でします。

もう一つ、気になるのは、一つだけDというステーションを設けたところで、Dが特殊すぎて、養殖場からの距離依存的に安定同位体比が変わるとは言いきれないという批判が来そうだということである。それに対応するには、養殖場から下流のDまで、いくつかのサンプリングステーション C_1, \dots, C_m を設けて、A, C_1, \dots, C_m , D の場所に依存する変動 (A, C_1, \dots, C_m , D、それ俺の平均値の分散) が、ランダムに起こると考えられる変動 (全分散から平均値の分散を差し引いた値) に比べて、どのくらい大きいのかを論ずれば良い。これがF 検定だ。その結果、場所による変動が十分大きいと考えられたら、その平均値が、場所の特性 (流れや干満差の特性、河川水と海水の入れ替わり量) によって、に応じてどのように違っているのかを議論していくことになる。

この研究は、実際に行われて、その結果は 2009 年に Fisheries Science 誌に掲載されています。興味があれば読んでください。単なる思いつきを研究に仕上げただけなので、単なる業績稼ぎとして国内の雑誌に簡単に書いたのですが、今から思うと、国際誌に書けばもっと注目されたと思います。

Kouetsu KON, Naoya KAWAKUBO, Jyun-Ichi AOKI, Prsert TOGNUNUI, Ken-Ichi Hayashizaki and Hisashi KUROKURA. (2009) Effect of shrimp farming organic waste on food availability of deposit feeder crabs in a mangrove estuary, based on stable isotopic analysis. Fisheries Science, 75(3) 715-722 今孝悦先生は国際水産開発学研究室で学位を取得され、現在、東京海洋大学海洋環境科学部門の准教授をされている、バリバリの生態学の研究者です。私たちの研究室の出身者ですから、社会科学的な研究にも理解があります。生態学は、社会学や経済学との間で、境界領域を共有する分野ですから、もっと、コラボレーションを考えても良いでしょう。林崎先生を除き、全員、当時の研究室のメンバーです。林崎先生は北里大学の先生です。この時は安定同位体分析の担当ですが、林崎先生の御専門は数理生態学です。面白いおじさんで、数学に強いので、何かと相談に乗ってくれます。

Google

