Data in Collective Motion

The Big Statistical Issues
What this talk is not…

- Ranting!
- Bayesian v Frequentist
- Reproducibility, p-hacking etc
**Summary**

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true.

**Factors that influence this problem and some corollaries thereof.**

**Modeling the Framework for False Positive Findings**

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a $p$-value less than 0.05. Research is not most appropriately represented and summarized by $p$-values, but, unfortunately, there is a widespread notion that medical research articles is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R + 1)$. The probability of a study finding a true relationship reflects the power $1 - \beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, $\alpha$. Assuming that $c$ relationships are being probed in the field, the expected values of the $2 \times 2$ table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significance.
Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; when there is greater flexibility in design, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prestige; and when more teams are involved in a scientific field in chasing statistical significance.

Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. For these reasons, the scientific literature, including much of what we read in leading medical journals, cannot be trusted, and researchers should not waste time hunting for true relationships in such fields.

Characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships.

It can be proven that most claimed research findings are false.

Research is not most appropriately represented and summarized by p-values, but, unfortunately, there is a widespread notion that medical research articles should be interpreted based only on formal statistical significance.
What this talk is not...

- Ranting!
- Bayesian v Frequentist
- Reproducibility, p-hacking etc

What this talk is...

Statistical inference issues that I repeatedly see in collective motion, and some (imperfect) solutions.
Data in Collective Motion

Some

The Big Statistical Issues
Problem 1: Emergence
\[ \mathbf{D}_t = \mathbf{D}_{t-1} + a\mathbf{A} + c\mathbf{C} + \varepsilon \]
Group level pattern matching

Alignment

Attraction


Group level pattern matching

**Alignment**

**Attraction**

Couzin *et al.*: *J. Theor. Biol.* 2002

Group level pattern matching

Alignment

Attraction


Group level pattern matching

Alignment

Attraction


Many ‘interaction rules’ papers
Katz et al., Inferring the structure and dynamics of interactions in schooling fish. *PNAS* 2011
Herbert-Read et al., Inferring the rules of interactions of shoaling fish. *PNAS* 2011
Problem 2: multiple tests
Learn one function, not many
Problem 3: time series
One decision

Four data points
Model decaying acceleration

Deciphering Interactions in Moving Animal Groups
Jacques Gautrais\textsuperscript{1,2,*}, Francesco Ginelli\textsuperscript{3,4,5}, Richard Fournier\textsuperscript{6,7}, Stéphane Blanco\textsuperscript{6,7}, Marc Soria\textsuperscript{8}, Hugues Chaté\textsuperscript{3}, Guy Theraulaz\textsuperscript{1,2}

Control for autocorrelation

Inferring the rules of interaction of shoaling fish
James E. Herbert-Read\textsuperscript{a,1,2}, Andrea Perna\textsuperscript{b,1}, Richard P. Mann\textsuperscript{b}, Timothy M. Schaar\textsuperscript{a}, David J. T. Sumpter\textsuperscript{b}, and Ashley J. W. Ward\textsuperscript{a,3}
• Control for autocorrelation
• Fit a single function, with spatial structure
• Use a neural network to define a flexible function space
• Separate different stimuli (bonus problem)
• Control for autocorrelation
• Fit a single function, with spatial structure
• Use a neural network to define a flexible function space
• Separate different stimuli

Cleaner, more powerful inference
Focus on change points

Strandburg-Peshkin et al. Current Biology 2013
Embed the time correlations with a latent space!

Mann et al. *PLoS Comp. Biol.* 2013
Problem 4: Additivity

\[ F(A \cup B) = F(A) + F(B) \]
Fish Ain’t Physics

Richard P Mann

UPPSALA UNIVERSITET
Nonpairwise interactions in three-fish shoals.

Yael Katz et al. PNAS 2011;108:18720-18725
$P_y = P(Y|\text{social info}) = \left(1 + s^{-\Delta n}\right)^{-1}$

$P_x = 1 - P_y = \left(1 + s^{\Delta n}\right)^{-1}; \quad \Delta n = n_y - n_x$
Problem 5: Emergence (again)
Bringing individual and collective behaviour together

Explain the direction changes
Local interactions
High $P(← | M)$

Low $P(← | M)$
No self-organisation
A non-Markovian interaction

Low probability of changing direction
A non-Markovian interaction

High probability of changing direction
A non-Markovian interaction

Still raised probability of changing direction
6. Responses are ambiguous

7. Expected responses mirror regression to the mean
7. Expected responses mirror regression to the mean

Return of agent 1 to equilibrium AFTER agent 2 leaves.
Conclusion: The era of naïve model fitting & validation is over. Time to level up