What is the replication crisis and why does it matter? A Bayesian perspective

Mark Andrews & Thom Baguley

@xmjandrews @seriousstats

Nottingham Trent University

July 7, 2016
Research Methods Festival 2016

What is the replication crisis?

- The false positive rate in psychology may be as high as 2 in 3.

What is the replication crisis?

- The false positive rate in psychology may be as high as 2 in 3.
- The Open Science Collaboration (2015) conducted replications of 100 psychology studies that were published in 2008:
  - While 97% of the original studies were statistically significant, only 36% of the replications were significant.
  - The mean effect size of the replications were less than half the magnitude of the originals.
  - Only 39% of the studies were subjectively rated as having replicated the original result.

What is the replication crisis?

- The false positive rate in psychology may be as high as 2 in 3.
- The Open Science Collaboration (2015) conducted replications of 100 psychology studies that were published in 2008:
  - While 97% of the original studies were statistically significant, only 36% of the replications were significant.
  - The mean effect size of the replications were less than half the magnitude of the originals.
  - Only 39% of the studies were subjectively rated as having replicated the original result.
- Similar results were reported in *Social Psychology* (2014): Of 31 replication attempts, 12 (39%) were successful, and 16 were failures.

What is the replication crisis?

- The false positive rate in psychology may be as high as 2 in 3.
- The Open Science Collaboration (2015) conducted replications of 100 psychology studies that were published in 2008:
  - While 97% of the original studies were statistically significant, only 36% of the replications were significant.
  - The mean effect size of the replications were less than half the magnitude of the originals.
  - Only 39% of the studies were subjectively rated as having replicated the original result.
- Similar results were reported in *Social Psychology* (2014): Of 31 replication attempts, 12 (39%) were successful, and 16 were failures.
- …and in psychiatry (Tajika et al., 2015): Of 43 replication attempts, 16 (37%) were successful, 16 others directly contradicted the original, and 11 had substantially smaller effects.

A Simple Bayesian Model of False Discoveries

- What is the probability that $\mathrm{H}_0 = \mathrm{True}$ given that $\mathrm{p} \leqslant \alpha$?

[1] Based on Ioannidis (2005).

A Simple Bayesian Model of False Discoveries

- What is the probability that $\mathrm{H}_0 = \text{True}$ given that $\mathrm{p} \leqslant \alpha$?
- We know that $\mathrm{P}(\mathrm{p} \leqslant \alpha | \mathrm{H}_0) \triangleq \alpha$, but what about $\mathrm{P}(\mathrm{H}_0 | \mathrm{p} \leqslant \alpha)$?

[1] Based on Ioannidis (2005).

A Simple Bayesian Model of False Discoveries

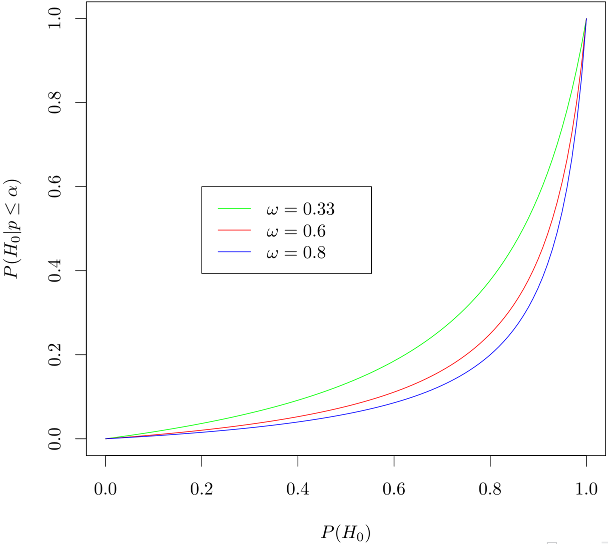
- What is the probability that $H_0 = \text{True}$ given that $p \leqslant \alpha$?
- We know that $P(p \leqslant \alpha | H_0) \triangleq \alpha$, but what about $P(H_0 | p \leqslant \alpha)$?

$$
\begin{aligned}
P(H_0 | p \leqslant \alpha) &= \frac{P(p \leqslant \alpha | H_0) P(H_0)}{P(p \leqslant \alpha | H_0) P(H_0) + P(p \leqslant \alpha | \neg H_0) P(\neg H_0)}, \\
&= \frac{\alpha(1-\lambda)}{\alpha(1-\lambda) + \omega\lambda}, \\
&= \frac{\alpha}{\alpha + \omega \frac{\lambda}{1-\lambda}}
\end{aligned}
$$

where α is the Type I error rate, ω is the power, and $\frac{\lambda}{1-\lambda}$ is the prior odds that $\neg H_0$ is true[1].

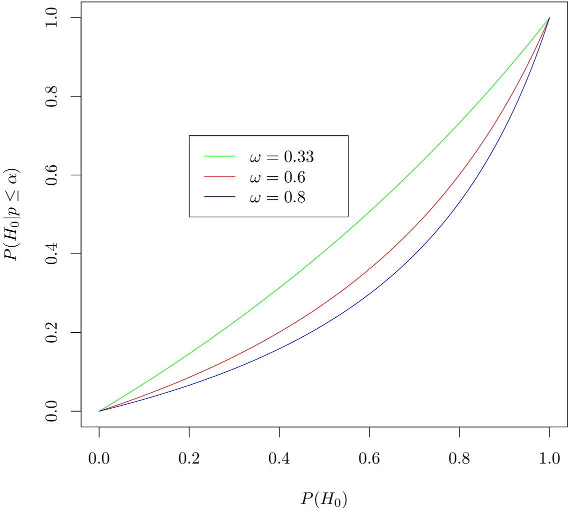
[1] Based on Ioannidis (2005).

Probability of false discovery



Probability of false discovery

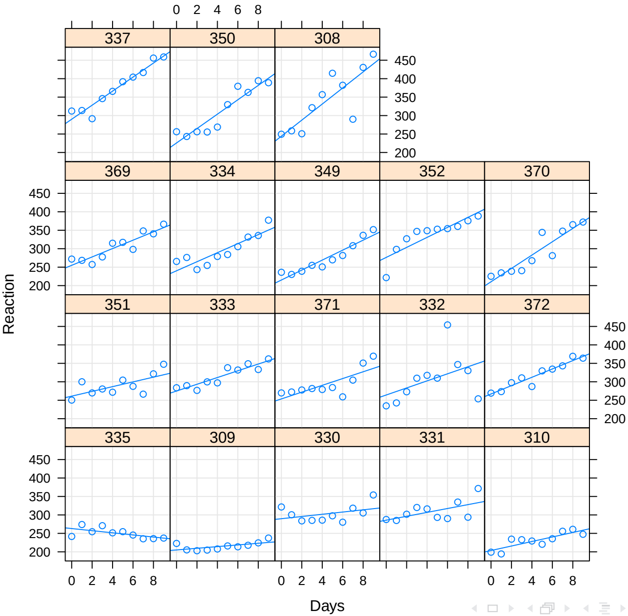
When there are $\mathsf{K} = 5$ *multiplicities*



Bayesian Meta-Analysis and Meta-Science

- A consequence of the replication crisis has been a demand for full-disclosure of scientific results.
- Modelling the (raw) data from many heterogeneous studies is effectively hierarchical modelling.
- Bayesian methods, and arguably *only* Bayesian methods, allow flexible modelling of such complex problems.

Bayesian Meta-Analysis and Meta-Science: Example



Bayesian Meta-Analysis and Meta-Science: Example

- We can model reaction time as a function of days without sleep with the following hierarchical model:

$$
\begin{aligned}
y_{ji} &\sim \alpha_j + \beta_j x_{ji} + \epsilon_{ji}, \\
\alpha_j &\sim N(a, \sigma_a^2), \quad \beta_j \sim N(b, \sigma_b^2)
\end{aligned}
$$

- We can model the same phenomenon across K different experiments with

$$
\begin{aligned}
y_{ji}^k &\sim \alpha_j^k + \beta_j^k x_{ji} + \epsilon_{ji}, \\
\alpha_j^k &\sim N(a^k, \sigma_a^2), \quad \beta_j \sim N(b^k, \sigma_b^2), \\
a^k &\sim N(a_0, \tau_a^2), \quad b^k \sim N(b_0, \tau_b^2).
\end{aligned}
$$

Bayesian Models of Multiple Simultaneous Inference

- At the heart of the replication crisis is the problem of multiplicities (QRPs, garden of the forking paths, etc).
- We can model multiple simultaneous inference with a hierarchical prior on null-effects. For example,

$$y_i = \sum_{k=1}^{K} \lambda_k \beta_k x_{ki} + \epsilon_i$$

where $\lambda_k \in \{0, 1\}$ is an indicator variable of non-null effects, and $\prod_{k=1}^{K} P(\lambda_k | \eta)$ is a hierarchical prior on non-null effects.

Bayesian Models of Multiple Simultaneous Inference: Example

- We can discover which of K coins are biased, after N flips each, using the following model:

$$
\begin{aligned}
y_k &\sim \text{dbinom}(p_k, N), \\
p_k &= \begin{cases} \theta_k & \text{if } \lambda_k = 1 \\ 0.5 & \text{if } \lambda_k = 0, \end{cases} \\
\theta_k &\sim \text{dbeta}(\alpha, \beta), \\
\lambda_k &\sim \text{dbern}(\pi)
\end{aligned}
$$

  where y_k is the observed number of Heads for coin k.
- Here, λ_k is a latent variable that indicates if the coin is biased or not.
- In simulations with $K = 100$ and $N = 100$, the false discovery rate ≈ 0.025, miss rate ≈ 0.075.

Conclusions

- What can Bayesian methods do for the replication crisis?
  - An understanding of discovery and the replication process.
  - More refined refined tools for research.
  - A more transparent research process.

References I

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med*, *2*(8), e124.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

Tajika, A., Ogawa, Y., Takeshima, N., Hayasaka, Y., & Furukawa, T. A. (2015). Replication and contradiction of highly cited research papers in psychiatry: 10-year follow-up. *The British Journal of Psychiatry*, *207*(4), 357–362.