How can the score test be consistent?

Natalie Karavarsamis[1]

Co-authors: Guillera-Arroita G.[1], Huggins R.[1], Morgan B.J.T.[2]
[1]School of Mathematics and Statistics, The University of Melbourne, Australia
[2]School of Mathematics, Statistics & Actuarial Science, University of Kent, UK

Outline

Comparing occupancies

A common question in ecology when studying occurrence of species is to compare two binomial proportions (ψ_1, ψ_2) under imperfect detection (p_1, p_2) as a way of comparing two occupancy samples or studies.

This leads to four parameters for estimation.

$H_0 : \psi_1 = \psi_2$

Available tests include

- Wald
- Score
- LRT

Comparing occupancies: Hypothesis tests

- Under H_0: Score, Wald & LRT are asymptotically equivalent
- Under H_1: tests are no longer equivalent; asymptotic theory may not hold
- Negative score test values may be produced for the score test using the observed information (*examined in paper*).
- Observed information is easy to compute numerically
- Closed form expressions for expected information do not always exist, especially for more complex models
- We propose a new modified rule based on the observed score test

Comparative Tests: Power

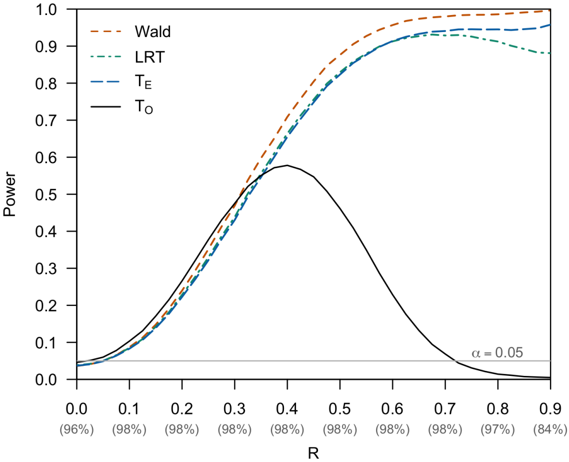

Figure 1: 50000 sims per effect size R ($\psi_2 = \psi_1(1 - R)$) where numerical optimization did not fail (shown as %).

Occupancy Model

Detections: independent Bernoulli trials.
Y_i are the number of detections over K ($y_i = 1, 2, \ldots, K$) visits at site i, $i = 1, \ldots, N$,

$$Pr(Y_i = 0) = 1 - \psi + \psi(1-p)^K$$
$$Pr(Y_i = y_i) = \psi p^{y_i}(1-p)^{K-y_i},$$

As the species is absent from some sites, the number of detections follows a zero-inflated binomial distribution (ZIB), with the level of zero-inflation set by $1 - \psi$.

$$L = \{\psi^{s_d} p^d (1-p)^{Ks_d - d}\}(1 - \psi\theta)^{N - s_d}, \theta = 1 - (1-p)^K$$

Note: no closed form expressions for the estimators (ie the score equations)

Score Test: Two-sample model

We wish to compare occupancy for two independent studies (samples).

$\boldsymbol{S}(\boldsymbol{\theta}) = (S_{11}, S_{12}, S_{21}, S_{22})^T$ unconstrained score function

$\boldsymbol{J}(\boldsymbol{\theta}) = \partial \boldsymbol{S}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}^T = \boldsymbol{S}'(\boldsymbol{\theta})$ observed information matrix

Observed Score Test Statistic under large-sample null distribution

$$\boldsymbol{T}_O(\boldsymbol{\theta}) = \boldsymbol{S}(\boldsymbol{\theta})^T \boldsymbol{J}(\boldsymbol{\theta})^{-1} \boldsymbol{S}(\boldsymbol{\theta}) \sim \chi_1^2,$$

replace $\boldsymbol{J}(\boldsymbol{\theta})$ with $E(\boldsymbol{J}(\boldsymbol{\theta}))$ for Expected Score Statistic $\boldsymbol{T}_E(\boldsymbol{\theta})$.

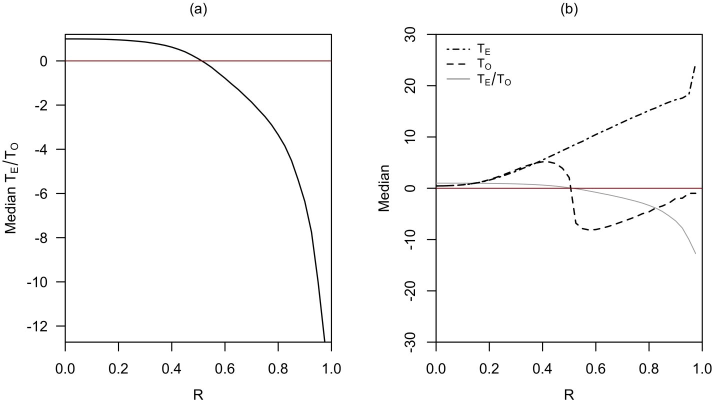

Figure 2: At $\psi_1 = \psi_2$, **the null hypothesis is true with effect size equal to zero, i.e.** $R = 0$. Then the score statistics are equal and their ratio is exactly equal to 1. At $R \approx 0.5$ half of the values of the observed score statistic are positive & half are negative.

The Score Test for Species Occupancy Model

Let

$\boldsymbol{\theta} = (\psi_1, p_1, \psi_2, p_2)^T$ model parameters,

$\boldsymbol{\theta}_T = (\psi_{1T}, p_{1T}, \psi_{2T}, p_{2T})^T$ true parameter values.

Consider

$H_0 : \psi_1 = \psi_2 = \psi,$

then let

$\boldsymbol{\theta}' = (\psi, p_1, p_2)^T$ model parameters under H_0

$\boldsymbol{S}_0(\boldsymbol{\theta}')$ score function under H_0.

$\boldsymbol{\theta}'_S$ is the restricted parameter subspace according to H_0

$E_{\boldsymbol{\theta}_T}(\boldsymbol{S}_0(\boldsymbol{\theta}'_S)) = 0$ is satisfied, and

$\widehat{\boldsymbol{\theta}}'_S$ is the MLE, and a solution of $\boldsymbol{S}_0(\boldsymbol{\theta}') = 0$

i.e. it maximises the log-likelihood subject to to the restricted subspace S.

The Score Test

The score test statistic defined in terms of the observed information is

$$\boldsymbol{T}_O(\widehat{\boldsymbol{\theta}}'_S) = \boldsymbol{S}(\boldsymbol{M}\widehat{\boldsymbol{\theta}}'_S)^T \boldsymbol{J}(\boldsymbol{M}\widehat{\boldsymbol{\theta}}'_S)^{-1} \boldsymbol{S}(\boldsymbol{M}\widehat{\boldsymbol{\theta}}'_S) \sim \chi^2_1$$

under H_0 asymptotically.

- Replace $\boldsymbol{J}(\boldsymbol{M}\widehat{\boldsymbol{\theta}}'_S)$ with $E_{\boldsymbol{\theta}_T}(\boldsymbol{M}\boldsymbol{J}(\boldsymbol{\theta}'_S))$ evaluated at $\boldsymbol{\theta}'_S = \widehat{\boldsymbol{\theta}}'_S$ to give the expected score test statistic $\boldsymbol{T}_E(\widehat{\boldsymbol{\theta}}'_S)$.
- As $\boldsymbol{\theta}_T$ is the true value $\widehat{\boldsymbol{\theta}}' \xrightarrow{P} \boldsymbol{\theta}'_S$, and $E_{\boldsymbol{\theta}_T}(\boldsymbol{J}(\boldsymbol{M}\boldsymbol{\theta}'_S))$ may be readily computed.
- This requires computing $\boldsymbol{\theta}'_S$ for a given $\boldsymbol{\theta}_T$.
- We examine these eigenvalues.

$$\boldsymbol{M} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

…derivations in paper (Arxiv.org)

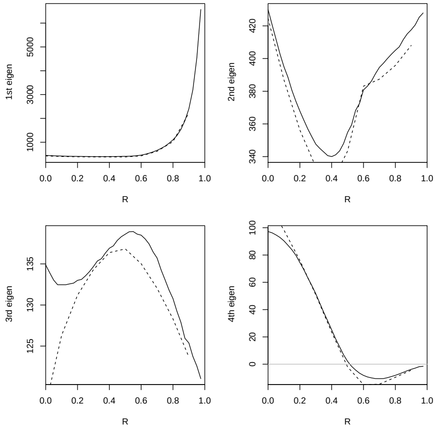

Figure 3: Eigenvalues for the observed information matrix, for different values of effect size R $\psi_1 = 0.8$. Solid lines are medians obtained from simulations (50000 at each value of R). Dashed lines are eigenvalues of $E_{\boldsymbol{\theta}_T}(\boldsymbol{J}(\boldsymbol{M}\boldsymbol{\theta}'_S))$.

The Score Test

When the null hypothesis is true,

$$\boldsymbol{J}(\widehat{\boldsymbol{\theta}}'_S) \longrightarrow \boldsymbol{I}(\boldsymbol{\theta}_T)$$

When H_0 is false, this is not so simple.

In our application the problem is that when

$$\boldsymbol{\theta}_T \neq \boldsymbol{M}\boldsymbol{\theta}'_S,$$

$E_{\boldsymbol{\theta}_T}(\boldsymbol{S}(\widehat{\boldsymbol{\theta}}'_S)) = f(\boldsymbol{\theta}_T, \boldsymbol{M}\boldsymbol{\theta}'_S)$ rather than
$E_{\boldsymbol{\theta}_T}(\boldsymbol{S}(\widehat{\boldsymbol{\theta}}'_S)) = f(\boldsymbol{\theta}_T)$ then
$E_{\boldsymbol{\theta}_T}\left(\boldsymbol{J}(\boldsymbol{M}\boldsymbol{\theta}'_S)\right)$ need not be positive definite.

- Ambiguous score function produces some positive and some negative eigenvalues of the observed information matrix. As a result, the observed score test statistic may be negative.

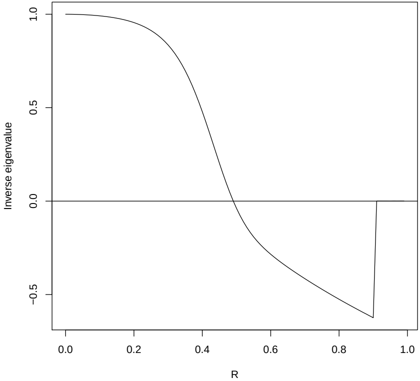

Figure 4: Inverse of the first eigenvalues of $\left(\left(E_{\boldsymbol{\theta}_T}\left(\boldsymbol{J}(\boldsymbol{M}\boldsymbol{\theta}'_S)\right)\right)^{-1} - \boldsymbol{M}\left(\boldsymbol{M}^T(E_{\boldsymbol{\theta}_T}\left(\boldsymbol{J}(\boldsymbol{M}\boldsymbol{\theta}'_S)\right)\boldsymbol{M}\right)^{-1}\boldsymbol{M}^T\right)\boldsymbol{\Sigma}$ as a function of effect size R. As in our earlier examination of $E_{\boldsymbol{\theta}_T}\left(\boldsymbol{S}(\boldsymbol{M}\boldsymbol{\theta}'_S)\right)$, we see that the eigenvalue becomes negative at $R \approx 0.5$. This confirms that the negative values of the score statistic are not just due to random variation.

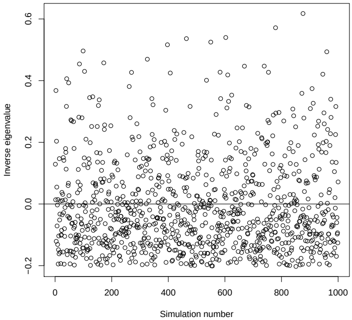

Figure 5: Inverse of the first eigenvalue of $\left((J(M\boldsymbol{\theta}'_S))^{-1} - M\left(M^T J(M\boldsymbol{\theta}'_S)M\right)^{-1} M^T \right)\boldsymbol{\Sigma}$ when $R = 0.6$ (1000 sims). Clearly, if there is only one nonzero eigenvalue and this is negative then the matrix must be negative definite. However, the values of the score statistic were observed in our simulations to be positive and negative. It is apparent that the eigenvalues for the observed information matrix can be negative or positive i.e. **random variation leads to the positive eigenvalues and hence positive values of the score statistics.**

Positive and negative scores

- The differences between T_O vs T_E are predominantly where the test based on the observed score statistic rejects the null hypothesis and that based on the expected score statistic does not.

- When we consider only those simulations where the observed score statistic is positive (T_O^+), we find there is good agreement between the expected (T_E) and observed (T_O^+) score test, i.e. both accept or reject the null hypothesis for a given dataset.
- As R increases, the number of datasets with positive tests n decreases substantially. We wish to increase n.

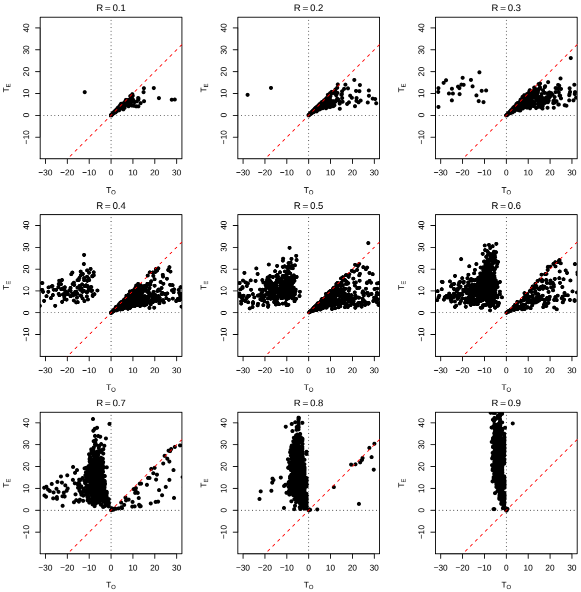

Figure 6: Agreement between observed (T_O) vs expected (T_E) score test statistic, for $\psi_1 = 0.8$.

Naive test: Observed Score Test

Naive use of the observed score test results in

- a test of **low power**, with
- **power decreasing as the alternative moves away from the null**, as we saw in the power plot.

The New Test

We were able to improve the power of the hypothesis test for occupancy data even when the information matrix contains negative values.

Our modified rule has

- power that is mostly greater to any other test and
- largely restores consistency.

The new test

The new test

Rejects the null hypothesis when the observed score statistic is larger than the usual chi-square cut-off or is negative.

Usual χ^2 rejection rule

$$T_O > \chi^2_{1,1-\alpha}$$

New rejection rule

$$T_O > \chi^2_{1,1-\alpha} \text{ or } \quad T_O < 0$$

New test is easy to use and inference is always possible.

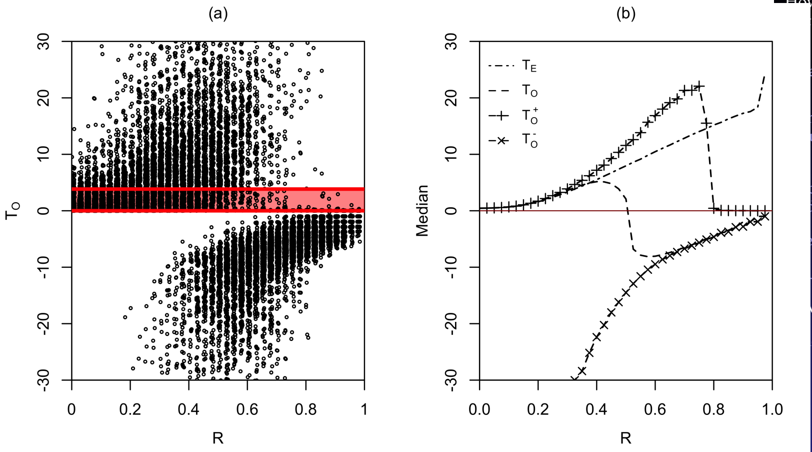

Figure 7: (a) Visual display of the new modified rejection rule for $\psi_1 = 0.8$. Power for each R is the proportion of simulations that lie outside the acceptance region.

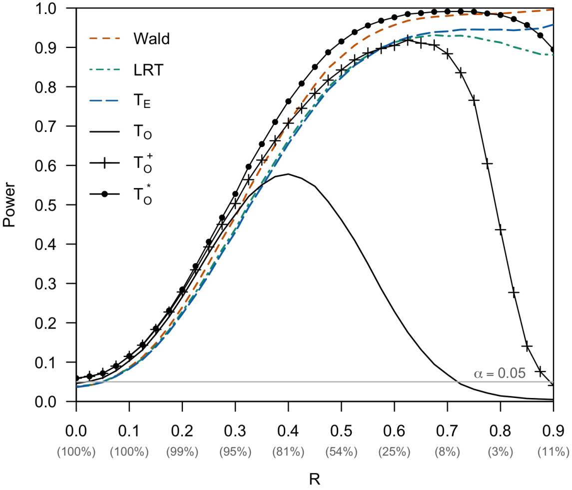

Figure 8: Power plot for $\psi_1 = 0.8$.

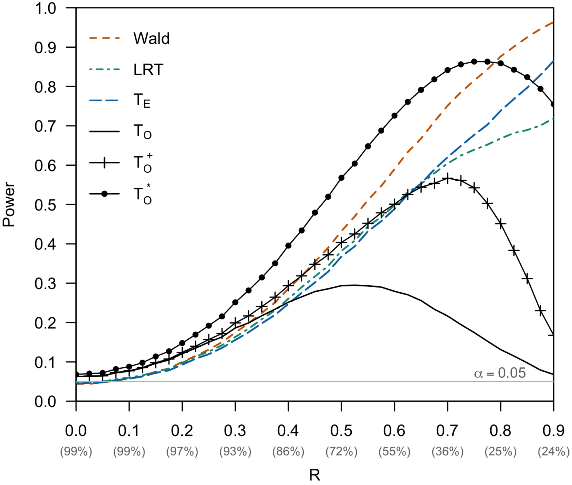

Figure 9: Power plot for scenario $\psi_1 = 0.4$

Background

Follows work done for zero-inflated Poisson by

- Freedman: How can the score test be inconsistent? (2007, *The American Statistician*, 61(4):291–295)
- Special section: Score Test oddities. Morgan BJT, Palmer KJ and Ridout MS (2007, *The American Statistician*, 61(4):291–295)

Summary

- At the unrestricted maximum, observed information will be usually positive definite.
- We compute observed information at $\hat{\theta}_S$, the parameter value maximising the log-likelihood over the null hypothesis, this is the restricted maximum.
- At a restricted max, the observed information can generate negative variance estimates - which makes inconsistency possible.

Problem

- The score test can be inconsistent because at the MLE under the null hypothesis, the observed information matrix produces negative variance estimates.
- The test can also be inconsistent if the expected likelihood equation has spurious (multiple) roots.

Problem

Freedman found

- expected model under the alternative is not always the same as under H_0 ie the asymptotics don't always work,
- this means an indefinite observed information matrix
- hence quadratic forms can be positive or negative
- this means there are negative eigenvalues
- that give positive or negative values in the observed info matrix
- that give negative score values...
- which means that the observed Score test can't be used...

Conclusions

Our new test

- is mostly the most powerful in our comparison to any other test
- is easy to use and inference is always possible
- restores consistency
- does not require lengthy algebra for obtaining analytic expressions for the expected information
- overcomes limitations when large sample assumptions fail and avoids contradictory results.
- works in practice when it is likely that an experiment may produce an indefinite information matrix

Thank you!

How can the score test be consistent?
Arxiv ID: 1805.05002