How to Fix Statistical Hypothesis Testing

Brian Dennis
Department of Fish and Wildlife Sciences
and
Department of Statistical Science
University of Idaho

Objective

Build model-comparison methods with improved statistical error properties to provide alternatives to Fisher-Neyman-Pearson hypothesis testing.

Approach

Combine R. Royall's concept of evidence with the model selection indexes descended from H. Akaike's pathbreaking work.

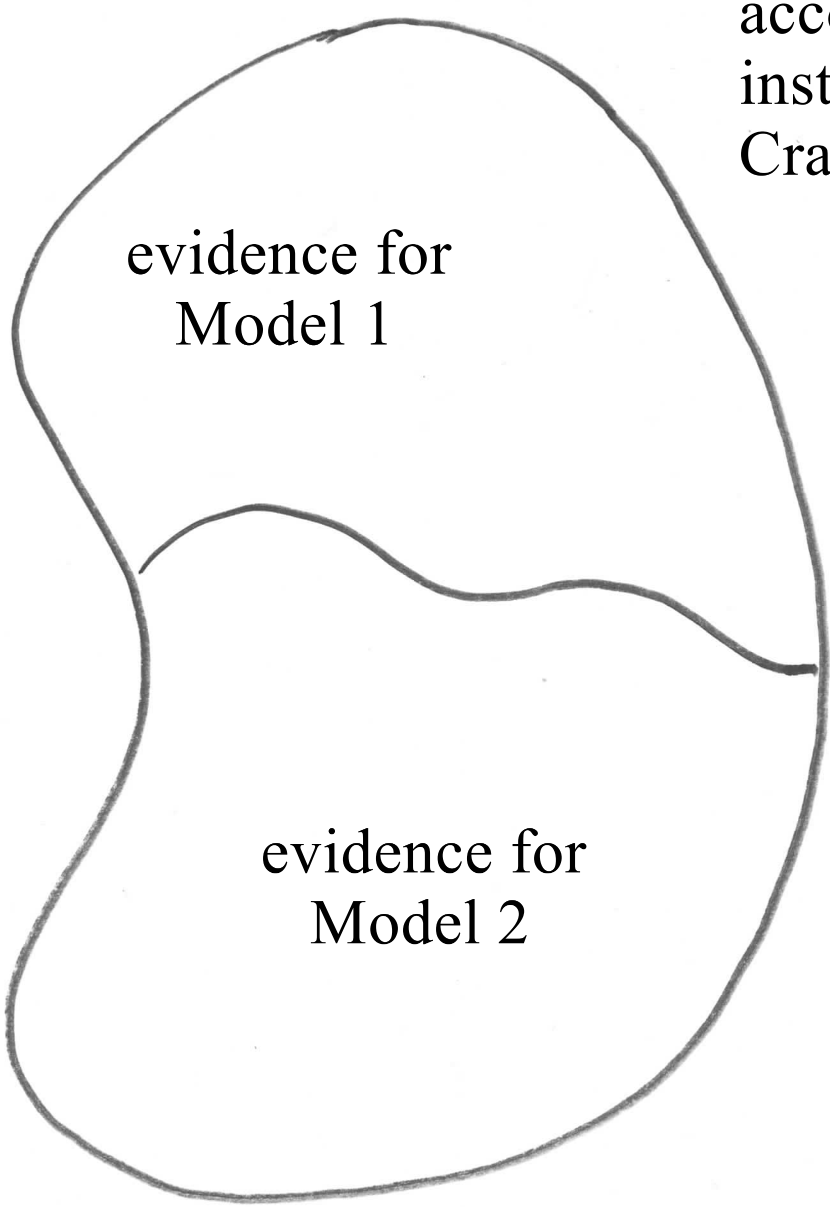
Collaborators

Mark Taper, José Ponciano, Subhash Lele

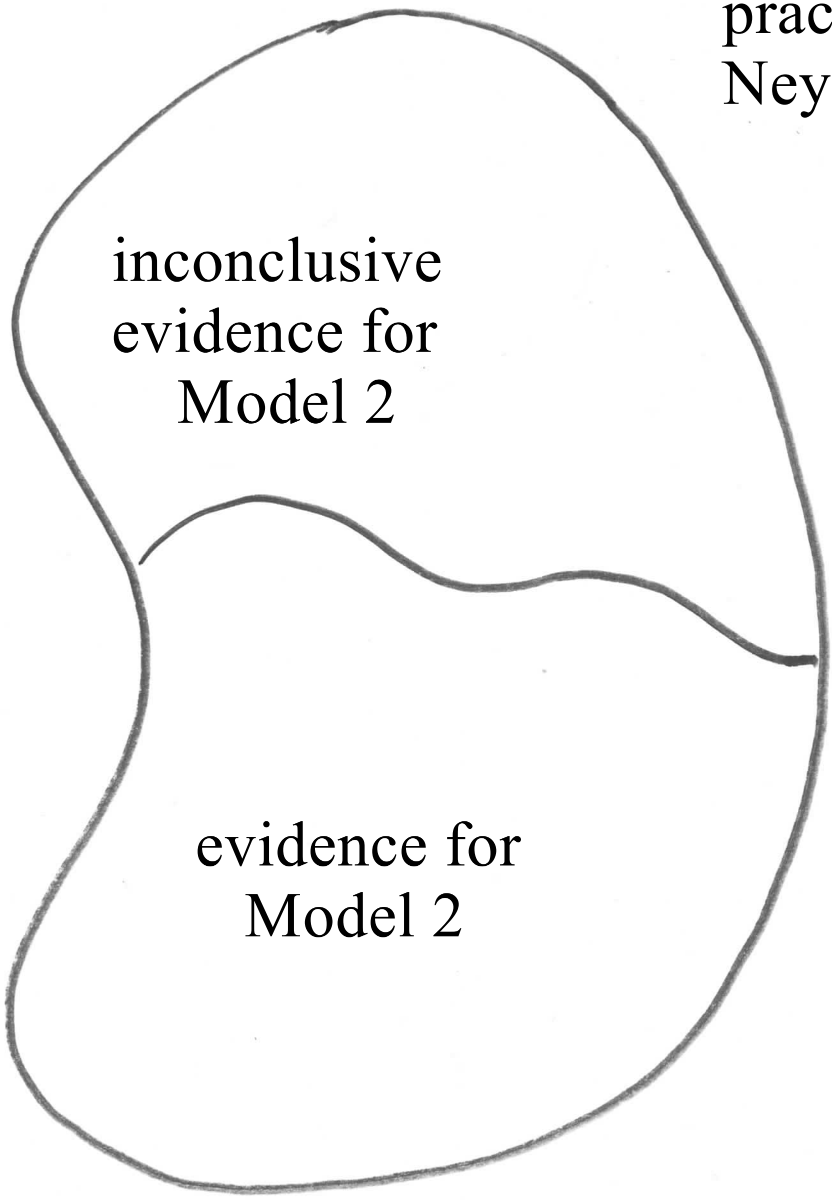
José Ponciano, Mark Taper, Subhash Lele

Sample space

Hypothesis testing, according to my math-stat instructor and to "Hogg and Craig," 1970s



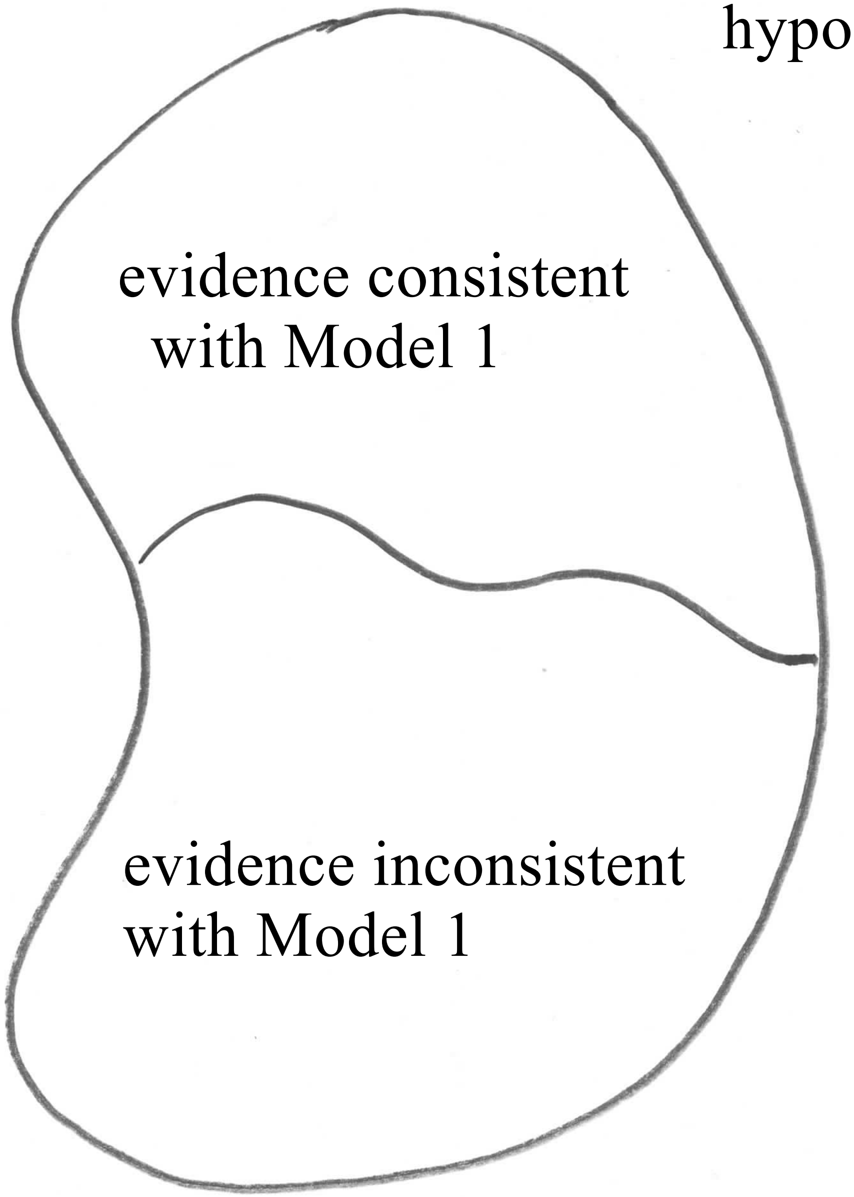
Hypothesis testing in practice, as emerged from Neyman and Pearson



inconclusive evidence for Model 2

evidence for Model 2

Fisher's original version of
hypothesis testing



evidence consistent
with Model 1

evidence inconsistent
with Model 1

Fisher (1920s)

H$_1$: Model 1 is a statistical distribution family (normal, binomial, etc.) with a parameter restricted (ex. $\mu = \mu_1$, $p = p_1$). No alternative hypothesis (other than *not* Model 1).

Model 1 provides a *likelihood function* and a statistic ($\widehat{\mu}$, $\widehat{p}$) that serves as an estimate of the parameter.

The plausibility of the data arising from Model 1 decreases as distance of the statistic from the restricted parameter value increases.

P value: the probability that the statistic would have a value as extreme as the observed value if the data generation mechanism were repeated.

Fisher suggested informaly that 0.05 would serve as a ordinary cutoff point for deciding about Model 1; if the value of P is smaller, the reasoned observer must regard the data under Model 1 as implausible. Model 1 is rejected.

Akin to what is called goodness of fit testing nowadays.

Neyman and Pearson (1930s)

To H$_1$, NP add H$_2$: Model 2, an alternative hypothesis.

Two types of errors: Type 1 (data came from Model 1 and wrongly points to Model 2), and Type 2 (vice versa).

Given data arise from Model 1, the probability of Type 1 error is the *size* of the test and is denoted α. Given data arise from Model 2, the probability of Type 2 error is denoted β. The *power* of the test is $1 - \beta$.

Neyman and Pearson sought to construct a statistical rule for deciding between Models 1 and 2 that would have good error properties.

They began by stripping the problem down to its essence; two models, both of which are completely specified (no unknown parameters, like binomial(40, $.5$) and binomial(40, $.75$)).

Neyman-Pearson setup:

$f_1(x)$: completely specified pdf for model 1 (H$_1$)

$f_2(x)$: completely specified pdf for model 2 (H$_2$)

x_1, x_2, ..., x_n: data (iid)

$$L_1 = f_1(x_1)f_1(x_2)\cdots f_1(x_n)$$

likelihood functions

$$L_2 = f_2(x_1)f_2(x_2)\cdots f_2(x_n)$$

$$\frac{L_1}{L_2} = \frac{f_1(x_1)f_1(x_2)\cdots f_1(x_n)}{f_2(x_1)f_2(x_2)\cdots f_2(x_n)}$$ likelihood ratio

Likelihood ratio test: decide on Model 1 if $L_1/L_2 > c$,

decide on Model 2 if $L_1/L_2 < c$,

where c is chosen so as to achieve a test of size α.

Neyman-Pearson Lemma (1933):

No other test of size α or less can have power greater than the power of the likelihood ratio test.

Subsequent history and consequences:

Wilks (1938) added result for dealing with unknown parameters (generalized likelihood ratio test: $G^2 = -2\,\log\left(\widehat{L}_1/\widehat{L}_2\right)$).

GLR test requires Model 1 to be nested in Model 2.

Asymmetric roles of Model 1 and Model 2 become baked into scientific practice; "fail to reject the null hypothesis" vs. "reject the null hypothesis" became the two possible decisions.

Methods for sorting among many contending models are jury-rigged from sequences of FNP pairwise tests (stepwise regression, multiple comparisons).

Data must arise from Model 1 or Model 2... "Type 3 error" of model misspecification could produce misleading results.

Studies are designed around $1 - \beta$ (power).

Concept of LR as Evidence (Royall 1997)

The LR has been proposed for decades as a measure of *evidence* for H$_1$ or H$_2$, principally by Hacking and Edwards. In the concept of evidence, the value of the LR itself is evidence, not an error rate that is pre-set, designed, or attained.

Royall took the NP setup and argued that reformulating the decision between Model 1 and Model 2 in terms of evidence improves the statistical properties of the decision.

Evidence-based test: decide on Model 1 (H$_1$) if L_1 is k times L_2; decide on Model 2 (H$_2$) if L_2 is k times L_1. Here k is a fixed threshold for the LR.

The approach produces a trichotomy of outcomes:

$k < \frac{L_1}{L_2}$: strong evidence for H_1

$1/k < \frac{L_1}{L_2} < k$: weak or inconclusive evidence

$\frac{L_1}{L_2} < 1/k$ (or $k < \frac{L_2}{L_1}$) : strong evidence for H_2

Here k is a fixed threshold for the LR value, not determined by error rates or sample size. Values of 8, 20, or 32 for k have been suggested.

The trichotomy of outcomes then leads to two types of errors, with two probabilities, given that data arise from H_1:

$$\mathsf{P}(\text{weak evidence} \mid \mathsf{H}_1) = \mathsf{P}\left(1/k < \tfrac{L_1}{L_2} < k \mid \mathsf{H}_1\right) \equiv W_1$$

$$\mathsf{P}(\text{misleading evidence} \mid \mathsf{H}_1) = \mathsf{P}\left(\tfrac{L_1}{L_2} < 1/k \mid \mathsf{H}_1\right) \equiv M_1$$

Similarly, given the data arise from H_2 :

$$\mathsf{P}(\text{weak evidence} \mid \mathsf{H}_2) = \mathsf{P}\left(1/k < \tfrac{L_1}{L_2} < k \mid \mathsf{H}_2\right) \equiv W_2$$

$$\mathsf{P}(\text{misleading evidence} \mid \mathsf{H}_2) = \mathsf{P}\left(k < \tfrac{L_1}{L_2} \mid \mathsf{H}_2\right) \equiv M_2$$

We can also define

$$\mathsf{P}(\text{strong evidence for } \mathsf{H}_i \mid \mathsf{H}_i) = \mathsf{P}\left(k < \tfrac{L_i}{L_j} \mid \mathsf{H}_i\right) \equiv V_i = 1 - W_i - M_i$$

Tasks at hand

A. Compare error properties of FNP testing and evidence testing, when models have been correctly specified (data arise from either $f_1(x)$ or $f_2(x)$).

B. Compare error properties of FNP testing and evidence testing, when models have been misspecified (data arise from a different pdf $g(x)$).

C. Extend the evidence concept to models with unknown parameters (will involve information-based indexes for model selection and the theory of maximum likelihood estimation when models are misspecified).

Studying properties of FNP vs evidence testing, when models correctly specified

$$K_{12} \;\equiv\; K(f_1, f_2) \;\equiv\; \mathrm{E}_1\left[\log\left(\frac{f_1(X)}{f_2(X)}\right)\right] \;=\; \sum\!\!\!\!\!\!\int f_1(x)\,\log\left(\frac{f_1(x)}{f_2(x)}\right)$$

$$K_{21} \;\equiv\; K(f_2, f_1) \;\equiv\; \mathrm{E}_2\left[\log\left(\frac{f_2(X)}{f_1(X)}\right)\right] \;=\; \sum\!\!\!\!\!\!\int f_2(x)\,\log\left(\frac{f_2(x)}{f_1(x)}\right)$$

Kullback-Leibler divergences

note: $\mathrm{E}_2\left[\log\left(\frac{f_1(X)}{f_2(X)}\right)\right] \;=\; \sum\!\!\!\!\!\!\int f_2(x)\,\log\left(\frac{f_1(x)}{f_2(x)}\right) \;=\; -\,\mathrm{K}_{21}$

$\frac{L_1}{L_2} \;=\; \frac{f_1(x_1)f_1(x_2)\cdots f_1(x_n)}{f_2(x_1)f_2(x_2)\cdots f_2(x_n)}$ likelihood ratio

$\log\left(\frac{L_1}{L_2}\right) \;=\; \sum_{i=1}^{n} \log\left(\frac{f_1(x_i)}{f_2(x_i)}\right)$ log-likelihood ratio

Note: taken as a random variable, the log-likelihood is a sum of iid random variables,

$$\log\left(\frac{L_1}{L_2}\right) = \sum_{i=1}^{n} \log\left(\frac{f_1(X_i)}{f_2(X_i)}\right) ,$$

so the CLT applies. Let

$$\begin{aligned}
{\sigma_1}^2 &= \mathsf{V}_1\left[\log\left(\frac{f_1(X)}{f_2(X)}\right)\right] = \mathsf{E}_1\left[\left(\log\left(\frac{f_1(X)}{f_2(X)}\right)\right)^2\right] - {K_{12}}^2 \\
&= \sum\!\!\!\!\!\!\int f_1(x) \left(\log\left(\frac{f_1(x)}{f_2(x)}\right)\right)^2 - {K_{12}}^2
\end{aligned}$$

$$\begin{aligned}
{\sigma_2}^2 &= \mathsf{V}_2\left[\log\left(\frac{f_1(X)}{f_2(X)}\right)\right] = \mathsf{E}_2\left[\left(\log\left(\frac{f_1(X)}{f_2(X)}\right)\right)^2\right] - {K_{21}}^2 \\
&= \sum\!\!\!\!\!\!\int f_2(x) \left(\log\left(\frac{f_1(x)}{f_2(x)}\right)\right)^2 - {K_{21}}^2
\end{aligned}$$

If the data arise from H$_1$:

$$\frac{\sqrt{n}}{\sigma_1}\left[\frac{1}{n}\text{log}\left(\frac{L_1}{L_2}\right) - K_{12}\right] \xrightarrow{\text{d}} \text{normal}(0,1)$$

Then:

$$\text{log}\left(\frac{L_1}{L_2}\right) \xrightarrow{\text{asymp}} \text{normal}(nK_{12}, n{\sigma_1}^2)$$ (Brownian motion approximaion)

$$\frac{1}{n}\text{log}\left(\frac{L_1}{L_2}\right) \xrightarrow{\text{asymp}} \text{normal}(K_{12}, {\sigma_1}^2/n)$$ (Sample mean distribution)

Similar expressions if data arise from H$_2$.

These limiting distributions allow straightforward approximations for the error probabilities α, β, P, W_i, M_i, V_i.

A. Results for correctly specified models

FNP testing:

As a function of n, α is fixed, β decreases toward zero.

As a function of $K_{12} + K_{21}$ (effect size), β decreases toward zero.

The decision threshold c is a rapidly changing function of n.

Evidence testing:

W_i, M_i decrease toward zero when n is large and increasing.

M_i increasses with n at first to a maximum value, then decreases.

V_i increases monotonically with n and asymptotes at 1. Also, $V_i > M_i$.

P is a post-data probability of misleading evidence against H$_1$ and depends primarily on the properties of H$_1$.

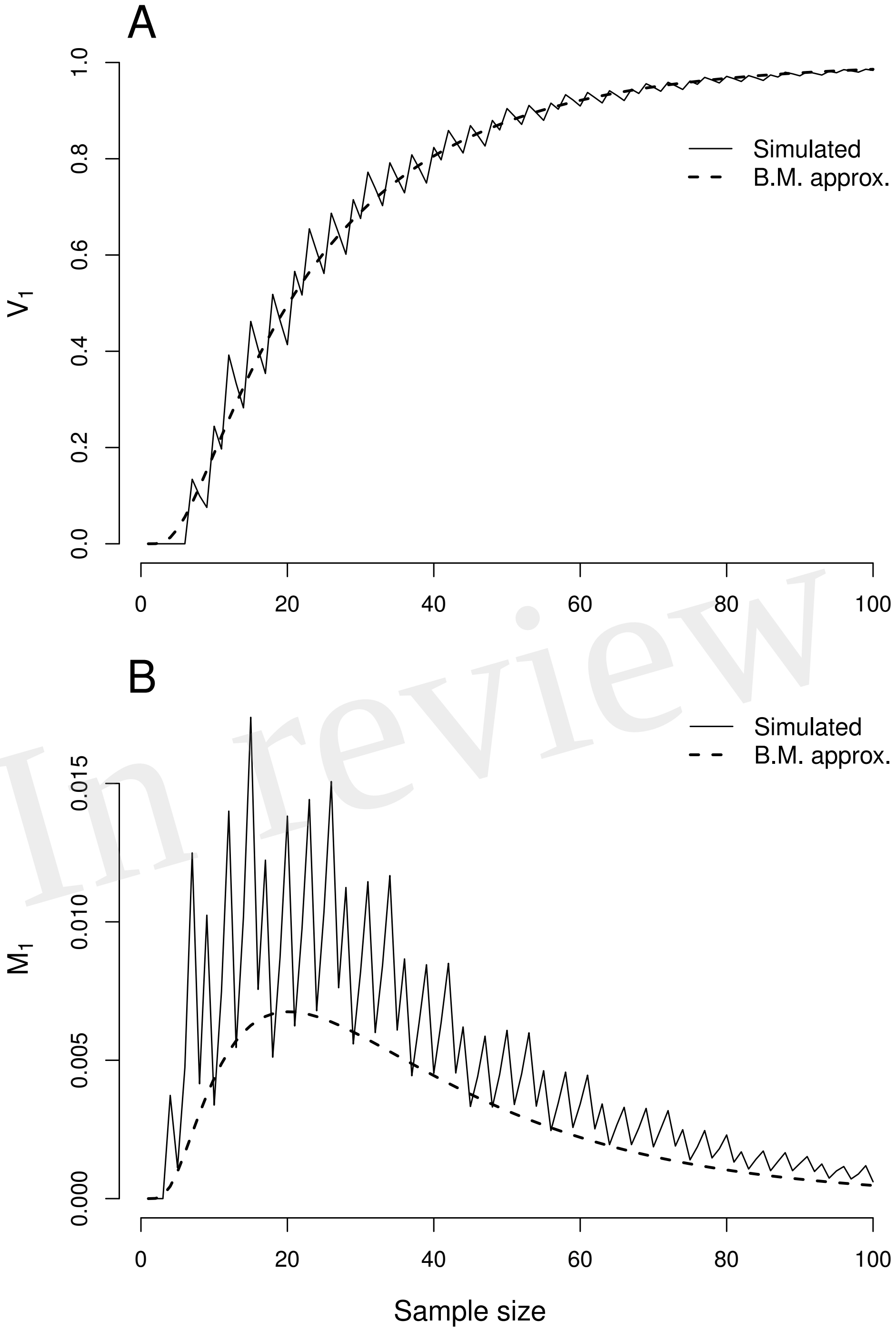


Figure 3. Evidence error probabilities for comparing two Bernoulli(p) distributions, with $p_1 = 0.75$ and $p_2 = 0.50$. Top panel **(A)**: simulated values (jagged curve) and values approximated under the Central Limit Theorem of the probability of strong evidence for model H_1, $V_1 = 1 - M_1 - W_1$. Bottom panel **(B)**: simulated values (jagged curve) and approximated values for the probability of misleading evidence M_1. Note that the scale of the bottom graph is one fifth of that of the top graph.

Studying testing properties when models are misspecified

To models $f_1(x)$ and $f_2(x)$ we add:

$g(x)$: The pdf for the "true" process that actually generates the data. And now,

$$\mathsf{E}_g\left[\log\left(\frac{f_1(X)}{f_2(X)}\right)\right] \;=\; \sum\!\!\!\!\!\!\int g(x)\log\left(\frac{f_1(x)}{f_2(x)}\right) \;=\; K(g,\,f_2) \,-\, K(g,\,f_1) \;\equiv\; \Delta K$$

Note that ΔK can be positive (f_1 closer to truth), negative (f_2 closer to truth), or zero (both models equally close). Define

$$\sigma^2 \;=\; \mathsf{V}_g\left[\log\left(\frac{f_1(X)}{f_2(X)}\right)\right] \;=\; \mathsf{E}_g\left[\left(\log\left(\frac{f_1(X)}{f_2(X)}\right)\right)^2\right] \,-\, (\Delta K)^2$$

$$=\; \sum\!\!\!\!\!\!\int g(x)\left(\log\left(\frac{f_1(X)}{f_2(X)}\right)\right)^2 \,-\, (\Delta K)^2$$

We then have

$$\frac{\sqrt{n}}{\sigma}\left[\frac{1}{n}\text{log}\left(\frac{L_1}{L_2}\right) - \Delta K\right] \xrightarrow{\text{d}} \text{normal}(0,1)$$

$$\text{log}\left(\frac{L_1}{L_2}\right) \xrightarrow{\text{asymp}} \text{normal}(n\Delta K, n\sigma^2)$$

$$\frac{1}{n}\text{log}\left(\frac{L_1}{L_2}\right) \xrightarrow{\text{asymp}} \text{normal}(\Delta K, \sigma^2/n)$$

We redefine "error" to be selecting the model "farthest" from $g(x)$ according to KL divergences. The various probabilities are denoted α', β', P', W_i', M_i', V_i'.

Now we use these CLT results to explore how errors behave under FNP and under evidence approaches.

B. Results under model misspecification

FNP testing:

α' is generally not equal to the advertised value of α and can be greater than or less than depending on the configuration of the models. It is easy to construct situations in which α' is a monotonic increasing function of n with an asymptote of one!

β' is generally not equal to β and can be greater than or less than depending on the configuration of the models. β' asymptotically decreases to zero as n becomes large.

P' is generally not equal to P and can be greater than or less than depending on the configuration of the models.

Evidence testing:

Lele's Lemma (2004)!! The probability V_i of strong evidence obtained with the LR evidence function cannot be exceeded by any other evidence function (taking misspecification into account and using KL divergence).

W_i', M_i', and V_i' are not in general equal to their counterparts from correct model specification. However, W_i' and M_i' eventually approach zero asymptotically as n becomes large, with M_i', increasing at first to a maximum before decreasing.

V_i' is a monotonic increasing function of n with an asymptote at 1.

$V_i' > M_i'$.

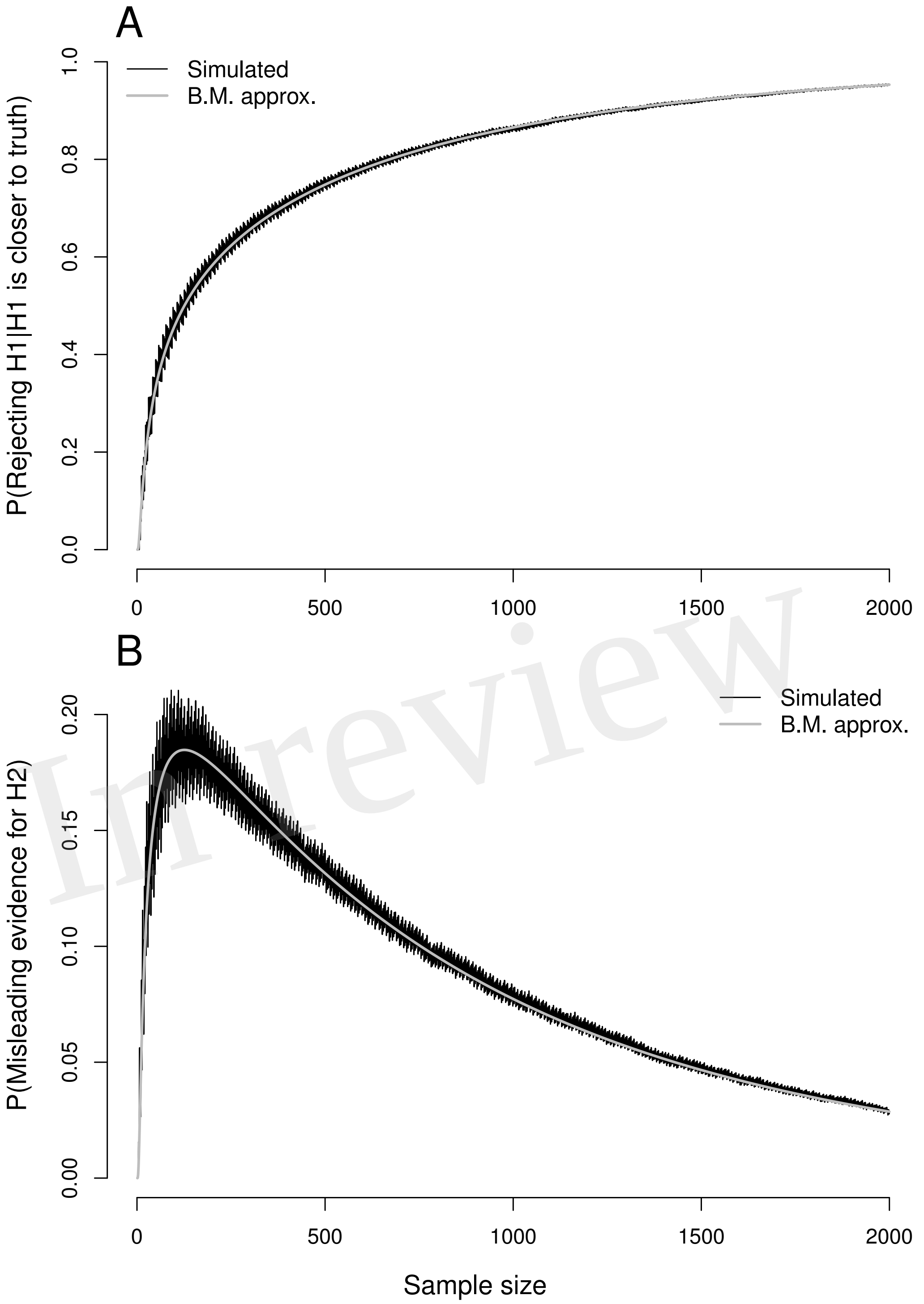


Figure 6. Evidence error probabilities for comparing two Bernoulli(p) distributions, with $p_1 = 0.75$ and $p_2 = 0.50$, when the true data-generating model is Bernoulli with $p = 0.65$. Top panel **(A)**: simulated values (jagged curve) and values approximated under the Central Limit Theorem of the probability (α') of rejecting model H$_1$ when it is closer than H$_2$ to the true model. Bottom panel **(B)**: simulated values (jagged curve) and approximated values for the probability (M_1') of misleading evidence for model H$_2$ when model H$_1$ is closer to the true data-generating process.

Studying testing when models have unknown parameters

Models with unknown parameters can have several different configurations.

FIGURE CAPTIONS

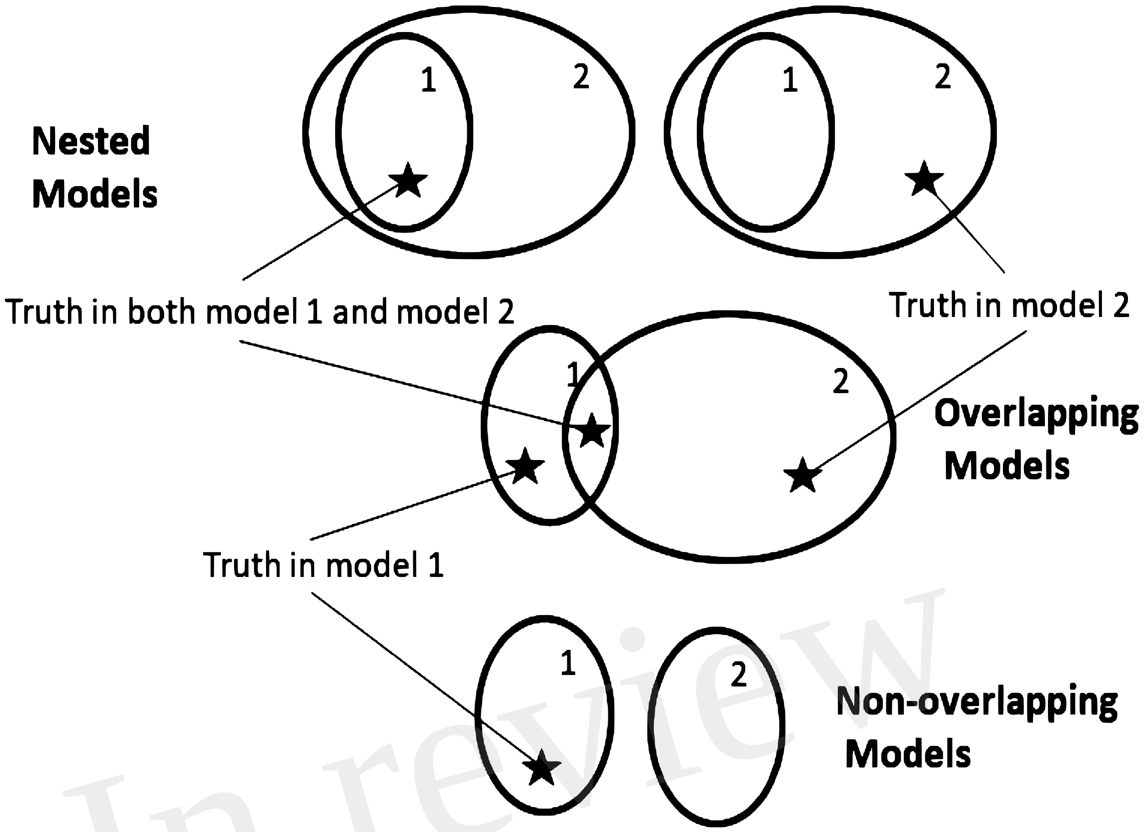


Figure 1. Model topologies when models are correctly specified. Regions represent parameter spaces. Star represents the true parameter value corresponding to the model that generated the data. Top: a nested configuration would occur, for example, in the case of two regression models if the first model had predictor variables R_1 and R_2 while the second had predictor variables R_1, R_2 and R_3. Middle: an overlapping configuration would occur if the first model had predictor variables R_1 and R_2 while the second had predictor variables R_2 and R_3. Three locations of truth are possible: truth in model 1, truth in model 2, and truth in both models 1 and 2. Bottom: an example of a nonoverlapping configuration is when the first model has predictor variables R_1 and R_2 while the second model has predictor variables R_3 and R_4

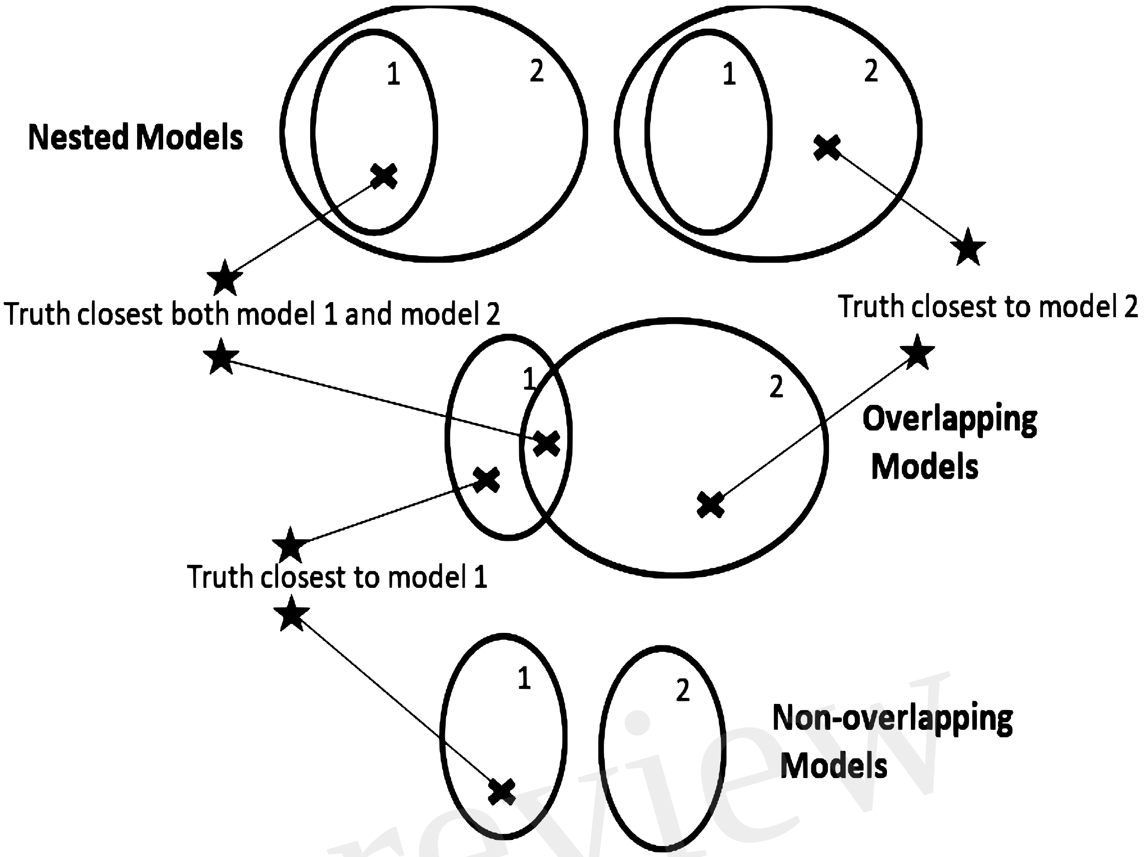


Figure 2. Model topologies when models are misspecified. Regions represent parameter spaces. Star represents the true model that generated the data. Exes represent the point in the parameter space covered by the model set closest to the true generating process.

Potential evidence functions can be built from information-theoretic indexes for model selection. Such evidence functions become functions of the generalized likelihood ratio statistic G^2.

$$\text{AIC}_i = -2\log\left(\widehat{L}_i\right) + 2r_i \qquad \text{(Akaike 1973)}$$

$$\Delta\text{AIC}_{12} = \text{AIC}_1 - \text{AIC}_2 = G^2 - 2\nu \qquad (\nu = r_2 - r_1)$$

$$\text{SIC}_i = -2\log\left(\widehat{L}_i\right) + r_i\log(n) \ \text{(Schwarz 1978)}$$

$$\Delta\text{SIC}_{12} = \text{SIC}_1 - \text{SIC}_2 = G^2 - \nu\log(n)$$

Asymptotic results from statistical theory, models correctly specified

Wilks (1938), Wald (1943), H_1 nested in H_2:

$G^2 \xrightarrow{\text{d}}$ chisquare(ν) under H_1.

$G^2 \xrightarrow{\text{asymp}}$ chisquare(ν, λ) under H_2, ($\lambda = nq$, q a Mahalanobis distance)

C1. Results for correctly specified models with unknown parameters, when models are nested or overlapping:

M_1 and W_1 for ΔAIC$_{12}$ ***do not*** go to zero as n becomes large. M_2 and W_2 do go to zero. AIC, when dealing with nested or overlapping models, has the error properties of FNP testing and not those of evidence testing.

M_1 and W_1 for ΔSIC$_{12}$ ***do*** go to zero as n becomes large. M_2 and W_2 go to zero as well. SIC, when dealing with nested or overlapping models, has the error properties of evidence testing.

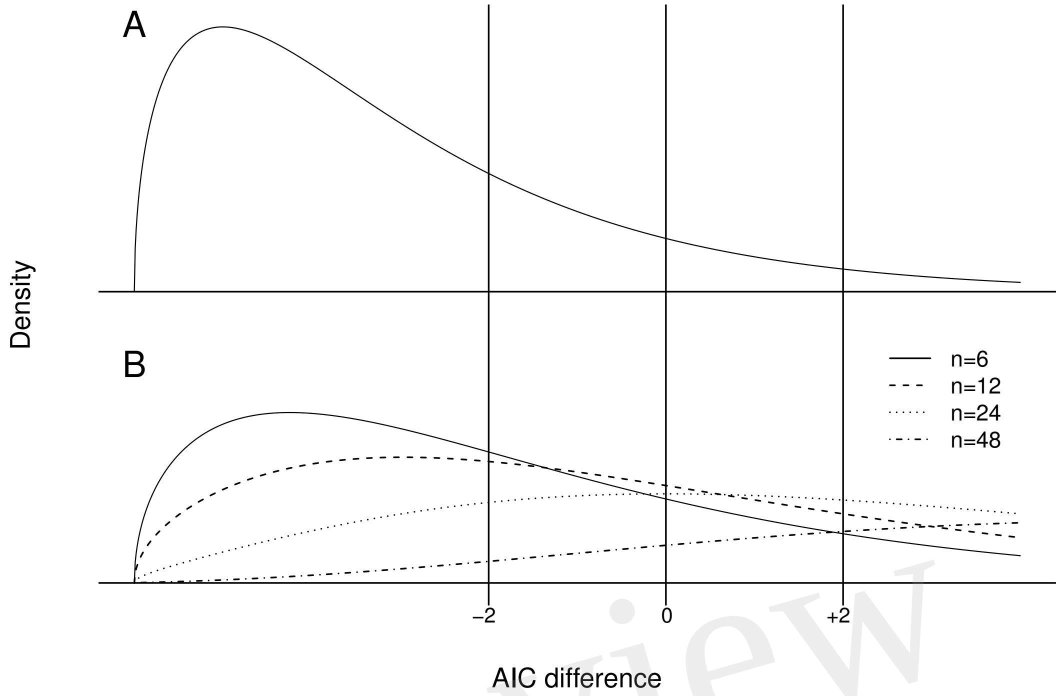


Figure 8. Top **(A)**: location-shifted chisquare distribution of the difference of AIC values, when data arise from model 1 nested within model 2. In this plot, the degrees of freedom for this distribution are equal to $\nu = 3$, and the shift to the left of 0 is equal $2\nu = 6$ (see equation 75 and text below it). This chisquare distribution is invariant to sample size. As a result, the areas under this distribution in the intervals $(-2, +2)$ and $(+2, \infty)$ corresponding to W_1 and M_1 respectively, are invariant to sample size. Bottom **(B)**: noncentral chisquare distribution of the difference of AIC values, when data arise from model 2 (but not model 1), plotted for different sample sizes. This distribution is also location-shifted but its noncentrality parameter λ, which determines both its mean and variance, is proportional to sample size. In this illustration, $\lambda = n(1/4)$. As a result, the areas under the intervals $(-2\nu, -2)$ and $(-2, +2)$ corresponding to the error probabilities M_2 and W_2 decrease as the sample size increases.

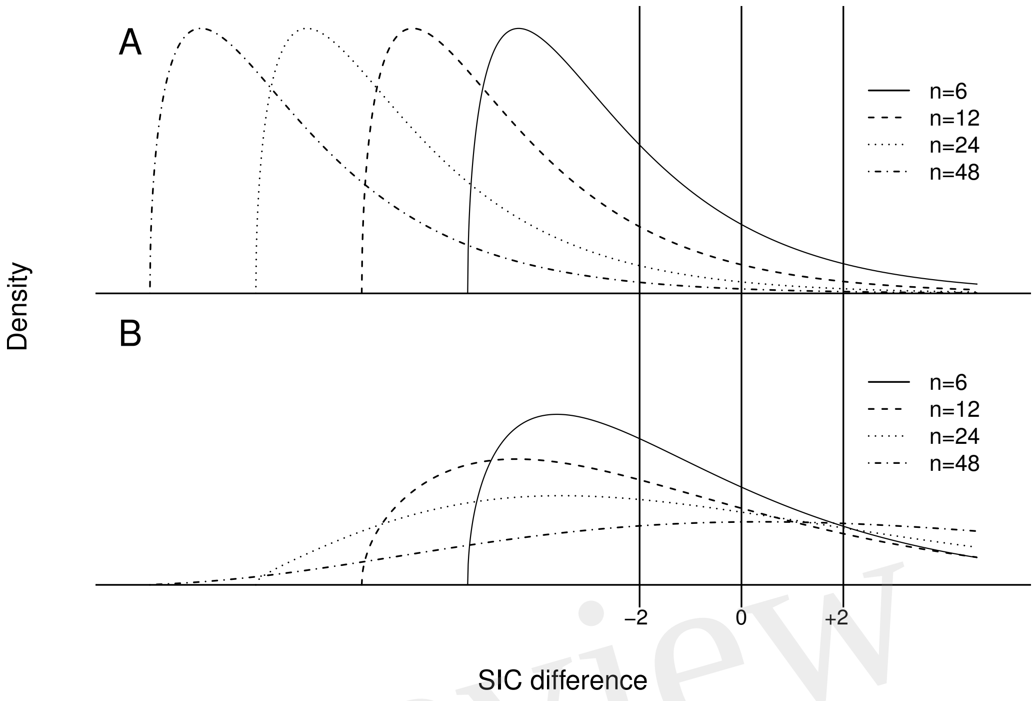


Figure 9. Top (**A**): Chisquare distribution of the difference of SIC values, when data arise from model 1 nested within model 2. The chisquare distribution is shifted left as sample size increases. Bottom **(B)**: noncentral chisquare distribution of the difference of SIC values, when data arise from model 2 (but not model 1), plotted for increasing sample sizes.

Asymptotic results from statistical theory, models misspecified

White (1982), Nishii (1988), Vuong (1989)

ML estimate $\widehat{\theta} \underset{\longrightarrow}{\text{p}} \theta^*$, where θ^* is value of θ minimizing $K(g(x), f(x,\theta))$

θ^* is in nested or overlapping region:

$G^2 \quad \xrightarrow{\text{asymp}} \quad$ weighted sum of chisquare(1) distributions

θ_i^* for the closest model is in nonoverlapping region:

$G^2 \quad \xrightarrow{\text{asymp}} \quad$ normal($2n\Delta K^*, 4n\sigma_*^2$) (Brownian motion!)

$$\Delta K^* = K(g(x), f_2(x,\theta_2^*)) - K(g(x), f_1(x,\theta_1^*))$$

$$\sigma_*^2 \;=\; \text{V}_g\left\{\text{log}\left[\tfrac{f_1(X,\theta_1^*)}{f_2(X,\theta_2^*)}\right]\right\}$$

C2. Results for misspecified models with unknown parameters

θ^* for the closest model is in nested or overlapping region:

M_1' and W_1' for ΔAIC$_{12}$ ***do not*** go to zero as n becomes large. M_2' and W_2' do go to zero. AIC, when dealing with nested or overlapping models, does not have error properties of evidence testing.

M_1' and W_1' for ΔSIC$_{12}$ ***do*** go to zero as n becomes large. M_2' and W_2' go to zero as well. SIC, when dealing with nested or overlapping models, has the error properties of evidence testing.

θ_i^* for the closest model is in nonoverlapping region:

M_1' and W_1' for both ΔAIC$_{12}$ and ΔSIC$_{12}$ go to zero as n becomes large. M_2' and W_2' go to zero as well. Both AIC and SIC, when dealing with nonoverlapping models, have the error properties of evidence testing.

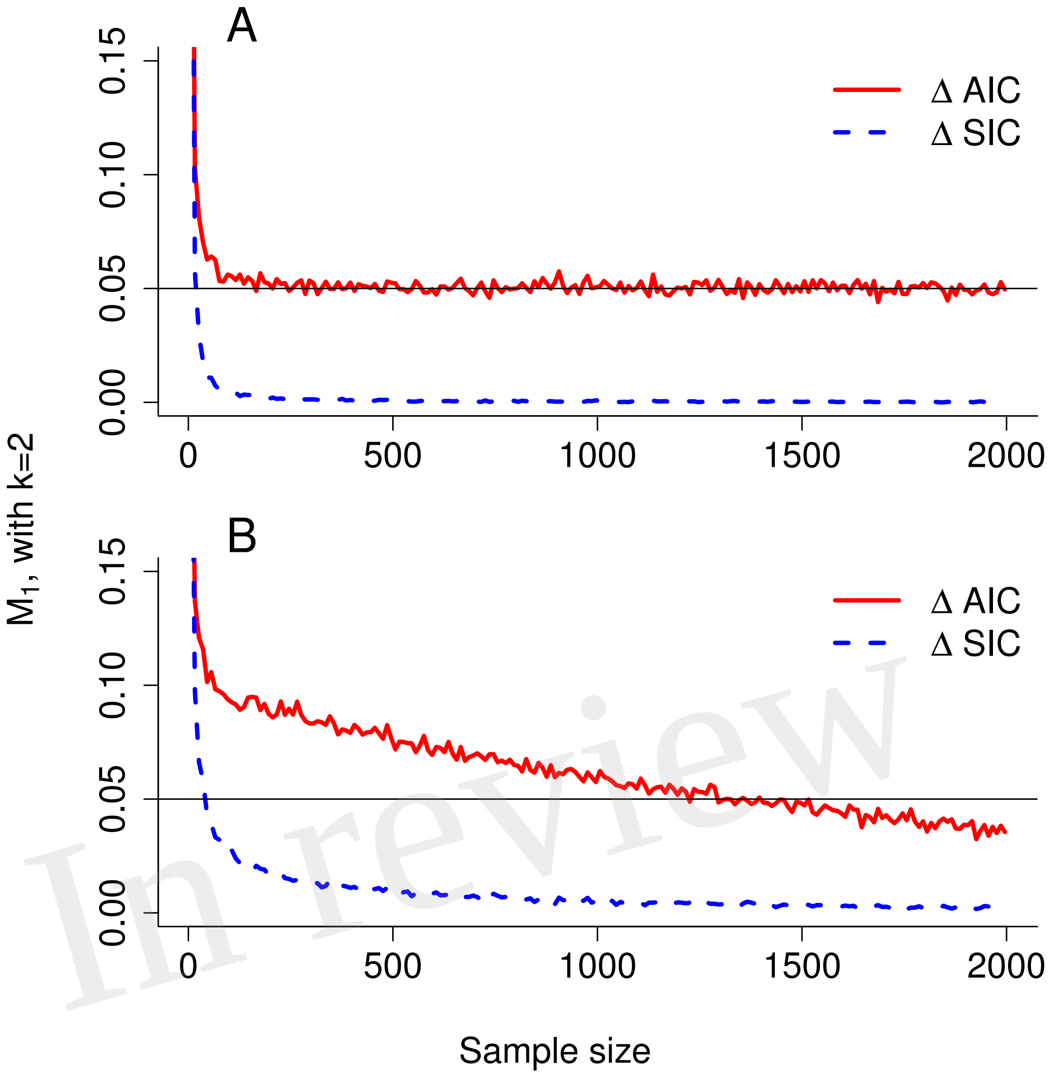


Figure 10. Simulation of Vuong's (1989) results for misspecified models. Top **(A)**: When $f_1\,(x,\ \theta_1{}^*)$ and $f_2\,(x,\ \theta_2{}^*)$ are the same model (either f_1 is nested within f_2, or f_1 overlaps f_2, and the best model is in the nested or overlapping region), then the asymptotic distribution of G^2 is a "weighted sum of chisquares" that does not depend on n. The error probabilities M_1 and W_1 do not decrease to 0 for ΔAIC_{12} but do decrease for ΔSIC_{12}. Bottom **(B)**: When the models are nested, overlapping, or nonoverlapping, but a nonoverlapping part of f_1 or f_2 is closer to truth, then G^2 has an asymptotic normal distribution with mean and variance that depend on the sample size and the error probabilities M_1 and W_1 decrease to 0 for both ΔAIC_{12} and ΔSIC_{12}. Details of these two settings in **(A)** and **(B)** are found in a fully commented R code.

How to do it

1. Use SIC, or any consistent model selection index based on likelihood ratio.

2. Choose k_1 and k_2, the decision thresholds. The decision rules will be:
   Strong evidence for Model 1 if $\Delta\text{SIC}_{12} \leq k_1$.
   Strong evidence for Model 2 if $k_2 \leq \Delta\text{SIC}_{12}$.
   Insufficient evidence if $k_1 < \Delta\text{SIC}_{12} < k_2$.

3. The thresholds k_1 and k_2 can be set by calculating (or simulating) the misleading evidence probabilities M_1 and M_2 for the applicable sample size n. The calculations or simulations are akin to power calculations. One procedure is to fit both models to the data (ML estimation), then simulate ΔSIC_{12} from each fitted model, estimating the two misleading evidence probabilities M_1 and M_2 for particular thresholds.

4. Rest well at night knowing that the evidence-based decsion will have some robustness to model misspecification.

TABLES

Table 1. A comparison of inferential characteristics between Fisherian significance testing (P-values *sensu stricto*), Neyman-Pearson hypotesis tests (including P-values for likelihood ratios) and evidential statistics.

Inferential characteristic	P-value	NP-test	Evidence
Equal status for Null and Alternatives	NA	No	**Yes**
Allows evidence for Null	No	No	**Yes**
Accommodates multiple models	No	Awkward	**Yes**
All error rates go to zero as sample size increases	No	No	**Yes**
Total error rate always decreases with increasing sample size	No	No	**Yes**
Can be used with non-nested models	NA	Not Standard	**Yes**
Evidence and error rates distinguished	No	No	**Yes**
Robust to model misspecification	Yes	No	**Yes**
Promotes exploration of new models	Yes	No	**Yes**

Akaike Guest House

Mar '71
17

TS

extension of the

[Information theory and an extended
maximum likelihood principle]

[Determination of the <u>number of factors</u> by
~~based on~~ a maximum likelihood principle

$$\underset{\theta}{\mathrm{Max}}\,\Pi f(x_i;\theta) = \Pi f(x_i;\theta^*)$$

sample

maximum number of allowable parameters

$$\int \frac{\log \Pi f(x_i;\theta)}{\log\ (\text{sample distribution density } f(x;\theta_0))}$$

$f(x;\theta_0)$ (sample distribution) dx

information

θ; which gives most near distribution to ~~to~~ the observed distribution of the variables

$$\int f(y;\theta_0)\left[\log f(y;\theta^*)\right] dy$$

$$\boxed{\ominus \int f(y;\theta_0) \log \frac{f(y;\theta^*)}{f(y;\theta_0)}\, dy}$$

expected <u>"distance"</u> of $f(y;\theta^*)$ from $f(y;\theta_0)$

(distance of θ^* from θ_0 as measured from the stand point of the distribution of y or as reflected in the distribution of y)

$\circledcirc$ { evaluation of θ^* as an estimate of θ_0 }

Max-L

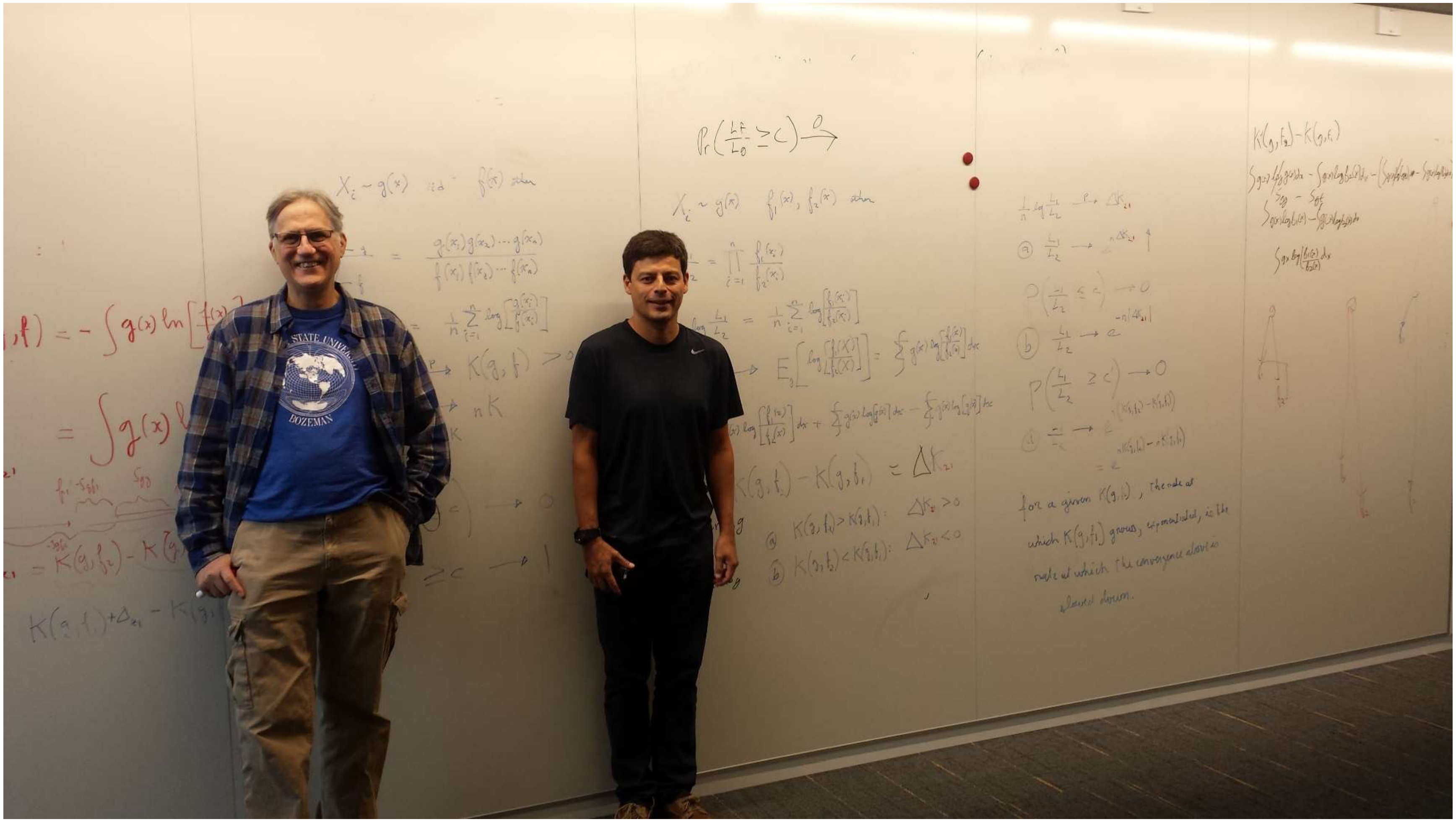
$$\left\{ \begin{array}{l} \Pi f(x_i;\theta) \\ \text{or } \log \Pi f(x_i;\theta) \end{array} \right\}$$

(-2) " log likelihood ratio "

$\sim \chi^2$ = stochastic distance function

● dimensionality (?)

→ [maximum allowable number of parameters
to specify the distribution


$P_r\left(\frac{L_f}{L_0} \geq c\right) \xrightarrow{} 0$
$X_i \sim g(x)$ iid $f(x)$ alter
$X_i \sim g(x)$ $f_1(x), f_2(x)$ alter
$K(g,f_1) - K(g,f_2) = \Delta K_{21}$
ⓐ $K(g,f_2) > K(g,f_1)$: $\Delta K_{21} > 0$
ⓑ $K(g,f_2) < K(g,f_1)$: $\Delta K_{21} < 0$
for a given $K(g,f_1)$, the rate at which $K(g,f_2)$ grows, exponentially, is the rate at which the convergence above is slowed down.
$K(g,f_2) - K(g,f_1)$
STATE UNIVE
BOZEMAN

brian@uidaho.edu

https://webpages.uidaho.edu/~brian/reprints/Dennis_et_al_errors_in_statistical_inference_under_misspecification_Front_Ecol_Evol_2019.pdf