Prevádzať p-hodnotu na Bayesov faktor?

[To translate p-value into Bayes factor?]

marian.grendar@savba.sk[1]

Robust 2012, Němčičky, Sep 9 - 14

[1] KM FPV UMB, BB; ÚM SAV BA; MÚ SAV BA.
Supported by VEGA 2/0038/12.

Outline

- Statistical evidence
- Sellke, Bayarri and Berger: p-value overstates evidence against point H_0
- p-value is not consistent measure of evidence
- SBB trafo of p-value into Bayes Factor
- Transform p-value into BF?

Intro

- Frequentist decision making: Neyman Pearson Wald
- Bayesian belief revision
- Evidence: what is the extent of support in data for a hypothesis H_0 relative to H_1

Setup, measure of evidence

r.v. $X \in \mathrm{R}^K$ with pmf/pdf $f_X(x \mid \theta)$, where $\theta \in \Theta \subseteq \mathrm{R}^L$.
partition of Θ: Θ_0, Θ_1
associate Θ_j with the hypothesis H_j, $j = 0, 1$.
$X_1^n \triangleq X_1, \ldots, X_n \sim f_X(x \mid \theta)$ random sample from $f_X(x \mid \theta)$.

Measure of evidence $\epsilon(H_0, H_1, X_1^n)$,

- in data X_1^n,
- for the hypothesis $H_0 : X_1^n \sim f_X(x \mid \theta)$ where $\theta \in \Theta_0$,
- relative to $H_1 : X_1^n \sim f_X(x \mid \theta)$ where $\theta \in \Theta_1$,

is a mapping $\epsilon(H_0, H_1, X_1^n) : \Theta_0 \times \Theta_1 \times (\mathrm{R}^K)^n \to \mathrm{R}$.

Measure of evidence *against* H_0, relative to H_1 denoted $\epsilon(\neg H_0, H_1, X_1^n)$.
Calibration. The partition of R that corr. to the extreme values of evidence that corr. to the strongest evidence is denoted S.

Measures of evidence

- Fisherian: p-value
- Likelihood-wallahs': ratio of likelihoods, extended ratio of likelihoods, ...
- Bayesian: Bayes factor, posterior odds, ratio of posterior modes, ...

Measures of evidence: Fisherian

The p-value $\pi(\neg H_0, \cdot, X_1^n) = \inf\{\alpha : T(X_1^n) \in R_\alpha\}$, where $T(\cdot)$ is a test statistic, α is the size of the test with the rejection region R_α for H_0.

Measures evidence in a data X_1^n, against a hypothesis H_0.
The smaller the p-value, the stronger the evidence against H_0 in the data.
The p-value in (0.01, 0.05) suggests strong, and smaller than 0.01 very strong evidence against H_0; i.e., $S = (0, 0.01)$

Measures of evidence: likelihood-wallahs'

The ratio of likelihoods $r_{01}(H_0, H_1, X_1^n) = f(X_1^n \,|\, H_0)/f(X_1^n \,|\, H_1)$.

Measures evidence in a data for a simple hypothesis H_0, rel. to a simple hypothesis H_1.

$r_{01} >$ app.30 suggests a very strong evidence for H_0 rel. to H_1; i.e., $S = [30, \infty)$.

Measures of evidence: Bayesian

Bayesian?

The Bayes Factor

$$BF_{01} = \int_{H_0} f(X_1^n \,|\, \theta) q(\theta) \, d\theta \Big/ \int_{H_1} f(X_1^n \,|\, \theta) q(\theta) \, d\theta,$$

where $q(\cdot)$ is the prior.

$BF_{01} > 150$, very strong evidence for H_0 rel. to H_1.

Sellke, Bayarri, Berger: hierarchical sampling

Sellke, Bayarri, Berger, '01

Ex.: yield (per hectare) of corn of sort D_l, $l = 1, 2, \ldots$.
Θ_0 corr. to 'mean yield is uninteresting',
Θ_1 corr. to 'mean yield is interesting'.
Experiment with corn D_l gives a random sample X_1^n.

Some sorts of corn give interesting mean yield, some give the uninteresting one.

i.e., some experimental data X_1^n come from H_0, other data sets are from H_1.

$$\text{Interest in: } \mathrm{P}(H_0 \,|\, \epsilon(\neg H_0, H_1, X_1^n) \text{ supports } H_1).$$

SBB: MC study of $\mathrm{P}\big(H_0\,|\,p\text{-}val \approx 0.05\big)$

Setup: $X_1^n \sim n(\theta = 0, 1)$.
Point $H_0 : \theta = 0$, point $H_1 : \theta = a$, $a > 0$.
π_0, proportion of data sets generated under H_0.
$\text{p-val} = 1 - \Phi(\sqrt{n}\,\overline{x})$.

$\mathrm{P}(H_0\,|\,\text{p-val} \approx 0.05)$; prob. that if p-val testifies strongly against H_0, then the data come from H_0.

a	n	π_0	$\mathrm{P}(H_0\,\vert\,\text{p-val} \in (0.04, 0.05))$
0.5	20	1/3	0.12
		0.5	0.25
		2/3	0.30

"... *p*-value near 0.05 provides at best weak evidence against H_0"

SBB setup: what about BF?

In the point-vs-point setting BF = likelihood ratio.

$\mathrm{P}(H_0 \,|\, BF_{10} > 150)$, analytically. Prob. that if BF testifies very strongly against H_0, then the data come from H_0.

a	n	π_0	$\mathrm{P}(H_0 \,|\, BF_{10} > 150)$
0.5	10	1/3	0.0015
		0.5	0.003
		2/3	0.006
		0.95	0.054
		0.99	0.23

Also BF can overstate evidence against H_0, though in more extreme case than p-val.

Consistency criterion

Data-sampling scheme:

1. First, θ is drawn from a pdf (or pmf) $p(\theta)$.
2. Given θ, a random sample X_1^n is drawn from $f_X(x \mid \theta)$.

We say that a measure of evidence $\epsilon(\neg H_0, H_1, X_1^n)$ against H_0, relative to H_1, is *consistent*, if

$$\lim_{n\to\infty} Pr(H_0 \mid \epsilon(\neg H_0, H_1, X_1^n) \in S) = 0.$$

The probability that θ is in Θ_0, given that the measure of evidence $\epsilon(\neg H_0, H_1, X_1^n)$ *very strongly* testifies against H_0, relative to H_1, should go to zero, as the sample size n goes beyond any limit.

p-val is not consistent, BF and LR are.

SBB: translation of p-val into BF

Under H_0, p-val is uniform. For a good test, under H_1 density $f(p)$ of p-val p should be decreasing in p. Take $H_1 : p \sim Beta(\xi, 1)$. Then, the lowest value $\underline{\mathrm{BF}}_{01}$ of BF_{01} over all priors $\pi(\xi)$ is:

$$\underline{\mathrm{BF}}_{01}(p) = -e\, p \log p.$$

The SBB trafo is meaningful for $p < 1/e$, where $\underline{\mathrm{BF}}_{01}(p)$ is increasing.

Ex. p-val $p = 0.05$ translates into $\underline{\mathrm{BF}}_{01}(p) = 0.407$, ie., almost none evidence against H_0.

SBB translation of p-val into BF: yes or no?

1) Since p-val is inconsistent and SBB translation does not depend on the sample size, the SBB-translated $BF(p)$ is inconsistent as well.

2) SBB trafo becomes asymptotically useless.

Ex. Take $n = 10^{12}$, and $p = 0.05$. SBB will translate it into $BF_{01} \geq 0.407$, whereas the true BF_{01} will with high probability give either very strong evidence for H_0 (if $X_1^n \sim f_0$), or for H_1 (if $X_1^n \sim f_1$).

3) The probab. that $p > 1/e$ converges to $(1 - 1/e)\pi_0$ as $n \to \infty$. Ex. for $\pi_0 = 0.5$ it is 0.32.

Good's translation of p-val into BF: yes or no?

Good: $BF_{01}^{G}(p) = \sqrt{n}\, p$.

Pros: Unlike SBB's, $BF_{01}^{G}(p)$ is consistent.
Cons: Good's BF cannot be greater than $\sqrt{n}$. Odd.

Summary

SBB demonstrate that p-val overstates evidence against H_0; and Bayes Factor does not. They propose a translation of p-val into BF.

BF can also overstate evidence against H_0. Unlike to p-val, however, to BF this happens asymptotically with zero probability. BF is consistent, p-val is not.

SBB translation of p-val into BF does not depend on n; the resulting measure of evidence is not consistent, and becomes asymptotically useless.

References

Good, I. J. (1992). The Bayes/Non-Bayes compromise: a brief review. *JASA*, 87(419):597-606.

g (2012). Is the p-value a good measure of evidence? Asymptotic consistency criteria, Stat. & Probab. Lett., 82, 1116-1119.

Sellke, T., Bayarri, M. J., and Berger, J. O. (2001). Calibration of p values for testing precise null hypotheses. *Amer. Statist.*, 55(1):62-71.

Shafer, G., Shen, A., Vereschagin, N., and Vovk, V. (2011). Test martingales, Bayes Factors and p-values. *Stat. Sci.*, 26(1):84-101.