Knowing the Signs:

Decision theory for significance tests

Ken Rice & Tyler Bonnett, University of Washington, Seattle, USA

Motivation

A brave new post p-value world? How might that look?

Must not have	Can live with	Must be
Spikes at $\theta = 0$	1D parameters	Bayesian
Conclusions $\theta = 0$	Parametric models	Decision Theoretic
	Frequentist properties	Connected to p's
		Scottish!

Scottish???

In 'Scots Law' there are *three* possible verdicts:

Verdict	Hypothesis test (Neyman-Pearson)	Significance test (Fisher)
Guilty	Reject H_0	Reject H_0
Not proven	no analog	No conclusion
Not guilty	Accept H_0	no analog

Decision theory for hypothesis tests

Loss functions deciding signs (is $\theta > 0$? $\theta < 0$?) are **very** limited.

Doing one-sided hypothesis tests we can **only** have:

		Decision	
		$d =$ Above	$d =$ Below
Loss when	$\theta > 0$	l_{TA}	l_{FB}
	$\theta < 0$	l_{FA}	l_{TB}

And with proper loss functions (see Phil Dawid's talk) this is wlog:

		Decision	
		$d =$ Above	$d =$ Below
Loss when	$\theta > 0$	0	α
	$\theta < 0$	$1 - \alpha$	0

... for some $0 \leq \alpha \leq 1$. The Bayes rule sets

$$d = \text{Above} \Longleftrightarrow \mathbb{P}[\theta < 0|\text{data}] < \alpha.$$

—acts like p's with large n, **but** no 'double the smallest tail'.

Decision theory for one-sided tests

Expressing one-sided **significance** tests as a decision:

		Decision	
		$d =$ Above	$\mathbf{d} =$**No Decision**
Loss when	$\theta > 0$	0	α
	$\theta < 0$	1	α

- 'Proper' loss fixes zero entry, and $0 \leq \alpha \leq 1$ ordering
- Assuming no decision is **equally bad**, regardless of truth

Different decision, same Bayes rule:

$$d = \text{Above} \Longleftrightarrow \mathbb{P}[\,\theta < 0 | \text{data}\,] < \alpha$$

—acts like one-sided p's with large n (*cf* Casella & Berger 1987)

Decision theory for two-sided tests

Using "no decision equally bad" and proper losses for two-sided decisions about $\theta's$ sign:

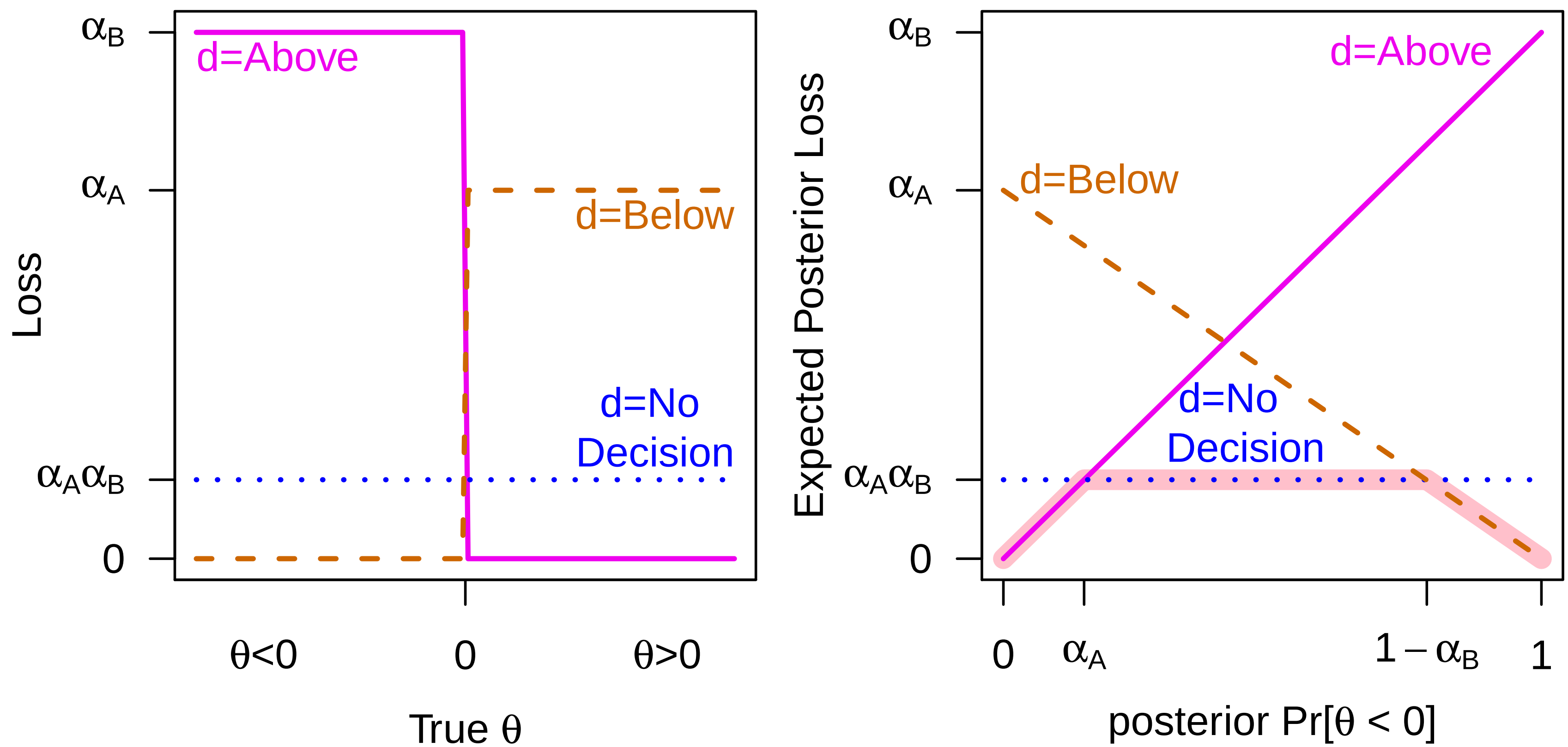
		Decision	
	$d =$ Above	d=No Decision	$d =$ Below
Loss when $\theta > 0$	0	$\alpha_A \alpha_B$	α_A
$\theta < 0$	α_B	$\alpha_A \alpha_B$	0
Bayes rule: do d iff	$\mathbb{P}[\theta < 0] < \alpha_A$	Otherwise	$\mathbb{P}[\theta > 0] < \alpha_B$

... and enforce $\alpha_A + \alpha_B \leq 1$ by insisting a randomized data-free d =Above/Below decision is worse than d =No Decision.

With symmetry, get a close Bayesian analog of two-sided tests:

		Decision	
	$d =$ Above	$d =$ No Decision	$d =$ Below
Loss when $\theta > 0$	0	α	2
$\theta < 0$	2	α	0
Bayes rule: do d iff	$\mathbb{P}[\theta < 0] < \alpha/2$	Otherwise	$\mathbb{P}[\theta > 0] < \alpha/2$

Decision theory for two-sided tests



- Two-sided significance tests are a close (large n) approximation of a Bayes rule **for choosing signs** – and up to 'proper' conditions, **no other losses/decisions are available**
- With symmetry, expected posterior loss = $\min\{P, \alpha\}$ for $P{=}2{\times}$minimum tail area – p-value tells us risk of sign-decision

How risky is it?

Frequentist expectation of this expected posterior loss, at fixed true θ;

$$\mathbb{E}[\min(P,\alpha);\theta] = \alpha\left(1 - \frac{\mathbb{P}[P<\alpha;\theta]}{\alpha}\mathbb{E}[\alpha - P|P<\alpha;\theta]\right)$$

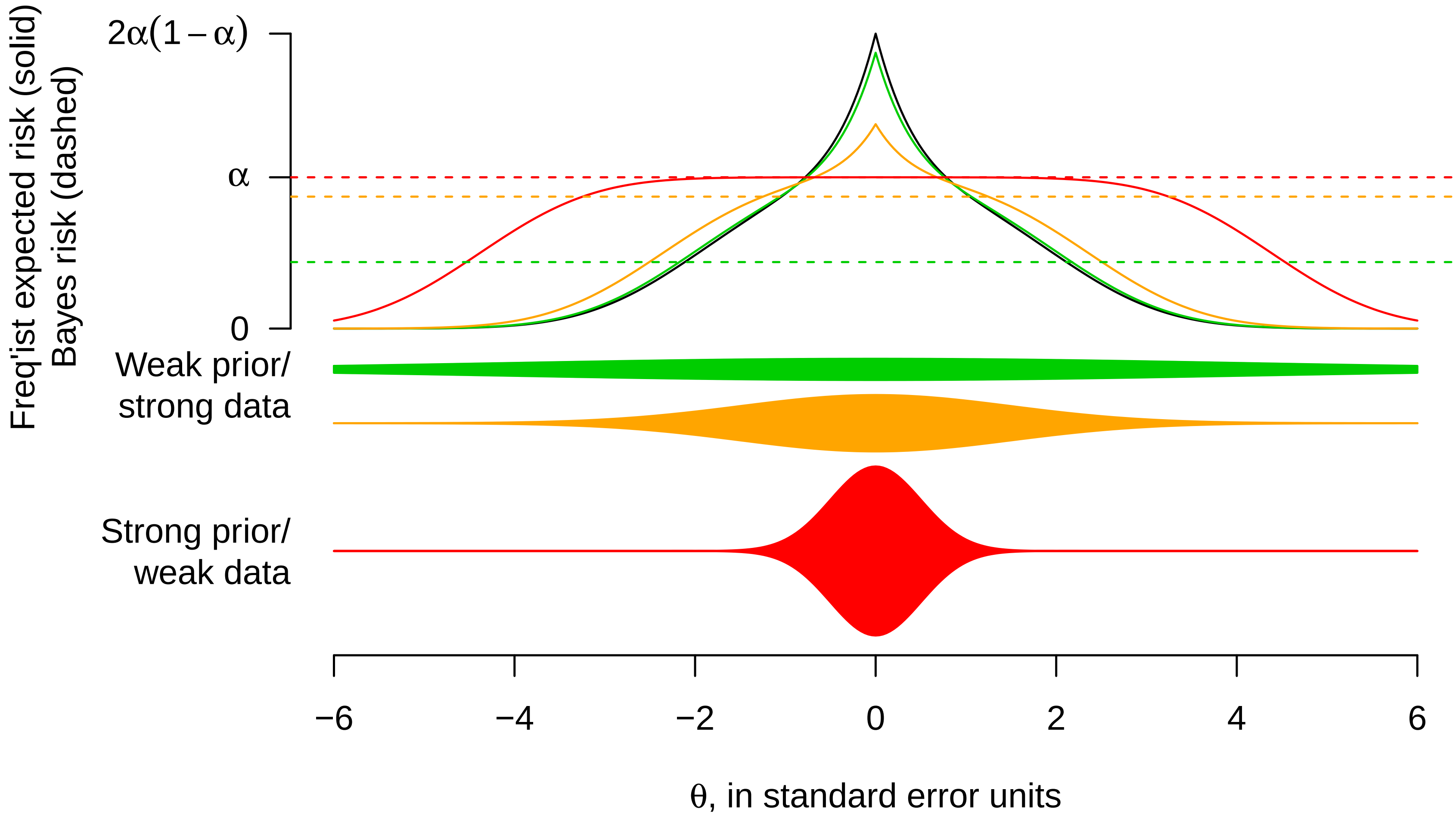
where $\mathbb{P}[P<\alpha;\theta]$ is the power (at large n).

- Ratio $\frac{\mathbb{P}[P<\alpha;\theta]}{\alpha}$ is Bayarri et al's (J Math Psych, 2016) *rejection ratio*, measuring 'evidentiary impact'

- Here (with a term in expectation of small p-values) it tells You how risky Your sign test decision would be, on average

How risky is it?

But full Bayes risk averages over datasets **and** prior. For various priors on Normal location problem ($N(\theta, \sigma^2)$, with σ^2 known):



Black curve gives risk of classic non-Bayes test

How risky is it?

- Fixing d =No Decision beats classic $\alpha = 0.05$ test (i.e. t-test) when it has ≤12% power ⇒ usual approach in crappy studies **is just futile!**
- Using full Bayes in those situations is a **little** better, but still **almost futile** – risk and Bayes risk very close to α
- Risk goes to zero as prior becomes improper – a reason to not use that prior!

Not shown:

- Similar behavior for other models/parameters
- Expected risk decomposes into two simple parts – frequentist replicates where P is bigger/smaller than observed P. Gives a Bayesian analog of *severity* (Mayo & Spanos)

Extensions: more nuanced decisions

Truth	Decision, d				
	Above	Suggest Above	No Decision	Suggest Below	Below
$\theta > 0$	l_{AA}	l_{Aa}	l_N	l_{Ab}	l_{AB}
$\theta < 0$	l_{BA}	l_{Ba}	l_N	l_{Bb}	l_{BB}

- Bayes rule determined by posterior tail area, again
- 'Proper' conditions on losses $\Longrightarrow$ means decision $A/a/N/b/B$ follows monotonically in left tail area
- Bayesian analog of recent Art Owen work – need more significant p-values to declare significance **and** sign of θ

Extensions: intervals

Written as a function, 2-sided loss with 'null' value θ_0 is

$$\alpha_B 1_{d=\text{Above}} 1_{\theta<\theta_0} + \alpha_A \alpha_B 1_{d=\text{No Decision}} + \alpha_A 1_{d=\text{Below}} 1_{\theta>\theta_0}$$

Making one decision for **each** possible null value θ_0, and adding the loss functions wrt non-negative measure on Θ, get

$$\alpha_B |\mathcal{A} \cap \{\theta_0 : \theta < \theta_0\}| \qquad + \alpha_A \alpha_B |\mathcal{N}| \qquad + \alpha_A |\mathcal{B} \cap \{\theta_0 : \theta > \theta_0\}|$$

for set-valued decisions $\mathcal{A}, \mathcal{B}, \mathcal{N}$.

- Bayes rule sets:
  - $\mathcal{A}$ to all θ_0 below low α_A quantile of posterior
  - $\mathcal{B}$ to all θ above high α_B quantile of posterior
  - $\mathcal{N}$ to the rest, i.e. the credible interval
- Fixing $\alpha_A + \alpha_B$ and choosing α_A gives centrality as a Bayes rule
- Measure on Θ says how to compare with other intervals

Extensions: tail area as decision

Significance loss trades-off Above/Below/No Decisions:

$$L(d,\theta) = 2 \times 1_{d=\text{Above}} 1_{\theta<0} + \alpha 1_{d=\text{No Decision}} + 2 \times 1_{d=\text{Below}} 1_{\theta>0}$$

A dual problem: decide the optimal price for **making** tradeoffs between these functions of θ:

$$L(s,a,\theta) = \frac{1}{\sqrt{a}}(2s1_{\theta<0} \qquad\qquad + a \qquad\qquad + 2(1-s)1_{\theta>0})$$

... for binary s and $0 \leq a \leq 1$. Note we **heavily** penalize tradeoffs that make No Decision cheap. Bayes rule sets:

- $s = 0/1$ depending if left/right tail is smaller
- $a = 2\times$minimum tail area

So, two-sided p-values (with direction of smallest tail) are approx Bayes, when deciding **how** to choose signs.

Extensions: multiple testing

For a single θ had loss: (one-sided for simplicity)

$$L(d,\theta) = 1_{d=\text{Above}} 1_{\theta<0} + \alpha 1_{d=\text{No Decision}}$$

For multiple θ_j and d_j, trading off total No Decision loss for a **single** wrong sign:

$$L(\boldsymbol{d},\boldsymbol{\theta}) = 1_{\cup_{j:d_j=\text{Above}}\{\theta_j<0\}} + \alpha\#\{d_j = \text{No Decision}\}$$

- A conservative approximation to the Bayes rule rejects null when $\mathbb{P}[\,\theta_j|\text{data}\,] < \alpha/m$, i.e. Bayesian Bonferroni correction
- Better (not conservative) – trade off mean sign-wrongness for total non-decisions, still get Bayesian Bonferroni
- Similar trade-offs provide Bayesian Benjamini-Hochberg algorithm (Lewis & Thayer 2009)

Extensions: Bayes Factors*

Bayes Factors compare posterior to prior – so not available from losses that use only θ. More Scottish inspiration...

Dolly the Sheep (1996–2003), first mammal cloned from an adult cell – at the University of Edinburgh, Roslin Institute

- We consider a *clone parameter* θ^*: same prior as θ, but **not** updated by data
- Decide if Sign(θ) $>$ Sign(θ^*)? Sign(θ) $<$ Sign(θ^*)? Or make no decision?

* ...Ba-a-a-ayes Factors?

Extensions: Bayes Factors*

To motivate Bayes Factors as one-sided significance test for $\theta > \theta^*$, **must** have loss

Truth		Decision, d	
		$d =$ Above	$d =$ No Decision
$\theta^* < 0$	$\theta < 0$	l_b	l_b
	$\theta > 0$	0	$\frac{1}{1+B}$
$\theta^* > 0$	$\theta < 0$	1	$\frac{1}{1+B}$
	$\theta > 0$	l_a	l_a
Bayes rule: do d iff		$\frac{\mathbb{P}[\theta>0]}{\mathbb{P}[\theta<0]}\frac{\mathbb{P}[\theta^*<0]}{\mathbb{P}[\theta^*>0]} > B$	$\frac{\mathbb{P}[\theta>0]}{\mathbb{P}[\theta<0]}\frac{\mathbb{P}[\theta^*<0]}{\mathbb{P}[\theta^*>0]} < B$

... for l_b, l_a and B all > 0.

- Bayes rule sets $d = A \iff \mathbb{P}[\,\theta < 0\,] = \frac{\mathbb{P}[\theta^*<0]}{B\mathbb{P}[\theta^*>0]+\mathbb{P}[\theta^*>0]}$
- Exactly the same as earlier significance tests, now with prior-dependent threshold $\alpha = \frac{\mathbb{P}[\theta^*<0]}{B\mathbb{P}[\theta^*>0]+\mathbb{P}[\theta^*>0]}$

Conclusions

A post p-value world? **Where knowing the signs is enough**, we can improve p-values **and much else** with Bayes **fairly easily**

- … and so instead focus on whether θ is *scientifically* relevant
- Need tools to convey that decisions are risky: plotting risk (and priors/posteriors) may help
- Knowing signs **is** enough in my applied work – see also Matthew Stephens' 'new deal' (2016, Biostatistics) on FDRs
- Please, please, don't claim p-values are evil/unBayesian