

LE-128/715

Segona prova

Víctor Peña

Universitat Politècnica de Catalunya

<http://vicpena.github.io/lector>

22 de desembre 2023

Continguts

1. Historial acadèmic i professional
2. Pla de treball
3. Lliçó

Historial acadèmic i professional

Educació

- 2008 - 2011 Diplomatura d'estadística
UPC
- 2011 - 2013 Màster en estadística i IO (MEIO)
UPC-UB
- 2013 - 2018 MS, PhD in Statistical Science
Duke University
Advisor: James O. Berger

Trajectòria professional

2018 - 2022 Assistant Professor in Statistics
Zicklin School of Business
The City University of New York

2022 - present Becari María Zambrano
UPC

Transferència de coneixement

- ▶ Consultoria: Takata (amb David Banks a Duke), SYMYB (amb Xavier Tort-Martorell)
- ▶ Formació a empreses: ORICA (amb Xavier Tort-Martorell i Lluís Marco)

Pla de treball

Docència

- ▶ Més de 8 anys d'experiència entre doctorat, assistant professor i becari
- ▶ Estudiantat divers: des d'estudiants d'escola de negocis fins a CFIS, des d'estudiants de primer de grau fins a doctorat
- ▶ Assignatures diverses: estadística bàsica, multivariant, R, bayesiana, disseny, modelització, inferència, etc.
- ▶ Docència en anglès i català

Recerca

Especialitat: teoria i mètodes bayesians

- ▶ Proves d'hipòtesi i selecció de variables
- ▶ Mètodes no paramètrics bayesians
- ▶ Sèries temporals
- ▶ Disseny i anàlisi d'experiments
- ▶ Altres: privacitat, optimització, computació i fonaments

Proves d'hipòtesi i selecció de variables

- ▶ Peña, V. & Barrientos, A.F. Differentially private methods for managing model uncertainty in linear regression models. **JMLR (Q1/Q2: CS-AI)**.
- ▶ Peña, V. & Barrientos, A.F. (2023) Differentially Private Hypothesis Testing with the Subsampled and Aggregated Randomized Response Mechanism. **Stat. Sinica (Q3)**.
- ▶ Mulder, J., Berger, J. O., Peña, V., & Bayarri, M. J. (2021). On the prevalence of information inconsistency in normal linear models. **TEST (Q2)**.
- ▶ Peña, V. & Berger J.O. (2020). Restricted type II maximum likelihood priors on regression coefficients. **Bayesian Analysis (Q1)**.

Mètodes no paramètrics bayesians

- ▶ Jauch, M., Barrientos, A. F., Peña, V. & Matteson, D. Mixture representations and Bayesian nonparametric inference for likelihood ratio ordered distributions. **Submitted to Bayesian Analysis (Q1).**
- ▶ Barrientos, A. F. & Peña, V. (2020). Bayesian bootstraps for massive datasets. **Bayesian Analysis (Q1).**

Sèries temporals

- ▶ Peña, V., & Irie, K. (2022). On the Relationship between Uhlig Extended and beta-Bartlett Processes. **Journal of Time Series Analysis (Q4)**.
- ▶ Investigació concurrent i futura amb Kaoru Irie (Universitat de Tokyo): models dinàmics multivariants per a matrius de covariàncies

Disseny i anàlisi d'experiments

- ▶ Attolini, C. S. O., Peña, V., & Rossell, D. (2015). Designing alternative splicing RNA-seq studies. Beyond generic guidelines. **Bioinformatics (Q1: Comp. Bio.)**.
- ▶ Investigació concurrent i futura amb Gonzalo García-Donato (Universidad de Castilla y la Mancha): anàlisi de dissenys factorials fraccionals

Altres

- ▶ Guo, Q., Barrientos, A.F. & Peña, V. Differentially Private Methods for Compositional Data. **Submitted to JCGS (Q1)**.
- ▶ Jauch, M. & Peña, V. (2016). Bayesian optimization with shape constraints. **NIPS Workshop on Bayesian Optimization**.
- ▶ Peña, V. & Berger, J. O. A note on recent criticisms to Birnbaum's theorem. *arXiv:1711.08093*.
- ▶ Investigació sobre mètodes computacionals per a la distribució GIG i estimació amb SURE amb Michael Jauch (FSU)

Finançament

- ▶ A CUNY: finançament intern pels estius i un projecte de recerca.
- ▶ Projecte del ministeri: “Métodos Bayesianos para la selección de variables en problemas de alta dimensionalidad y con datos perdidos” amb Gonzalo García-Donato (UCLM), Maria Eugenia Castellanos (URJC), Alicia Quirós (León), Stefano Cabras (UC3M) i Anabel Forte (UV)
- ▶ Presentarem proposta per la nova convocatòria de l'AEI (gener 2024)

Lliçó

p -values vs $P(H_0 \mid \text{data})$

- ▶ In introductory statistics classes, you learned that the p -value is not $P(H_0 \mid \text{data})$.
- ▶ In Bayesian statistics, we can compute $P(H_0 \mid \text{data})$.

Question

How similar is $P(H_0 \mid \text{data})$ to the p -value?

Testing a point null: normal mean

- ▶ We observe $X_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$ for $i \in \{1, 2, \dots, n\}$ with σ^2 known.
- ▶ We want to do the hypothesis test

$$H_0 : \mu = \mu_0 \qquad H_1 : \mu \neq \mu_0$$

for a fixed μ_0 .

Non-Bayesian solution: z -test

- ▶ From a non-Bayesian perspective, we can do a z -test.
- ▶ If Φ is the cdf of the $N(0, 1)$ distribution and $z = \sqrt{n}(\sum_{i=1}^n x_i/n - \mu_0)/\sigma$, the p -value is

$$2[1 - \Phi(|z|)] = 2[1 - \text{pnorm}(\text{abs}(z))].$$

- ▶ If we reject H_0 whenever the p -value is less than α , then

$$P_{H_0}(\text{reject } H_0) = \alpha.$$

Bayesian solution

- ▶ All unknowns have probability distributions.
- ▶ Need to specify $P(H_0)$ and $P(H_1)$.
- ▶ Under H_0 , we know that $\mu = \mu_0$.
- ▶ Under H_1 , we know that $\mu \neq \mu_0$, but we don't know the value of μ exactly: we need to specify a prior $f(\mu \mid H_1)$.

Bayes theorem

Applying Bayes theorem, we can find

$$P(H_0 \mid x_{1:n}) = \frac{P(H_0)f(x_{1:n} \mid H_0)}{P(H_0)f(x_{1:n} \mid H_0) + P(H_1)f(x_{1:n} \mid H_1)}$$

$$f(x_{1:n} \mid H_0) = \prod_{i=1}^n N(x_i \mid \mu_0, \sigma^2)$$

$$f(x_{1:n} \mid H_1) = \int_{\mathbb{R}} \underbrace{\prod_{i=1}^n N(x_i \mid \mu, \sigma^2)}_{f(x_{1:n}, \mu \mid H_1)} f(\mu \mid H_1) \, d\mu.$$

How to choose priors?

- ▶ How to choose $P(H_0)$ and $P(H_1)$?
 - ▶ In the absence of prior information, we can choose $P(H_0) = P(H_1) = 1/2$.

- ▶ How to choose $f(\mu | H_1)$?

- ▶ Use a default prior: for example,

$$\mu | H_1 \sim N(\mu_0, \sigma^2) \quad (\text{unit information prior})$$

- ▶ We can also consider robust Bayes.

Robust Bayes

- ▶ Instead of considering one prior, consider a class of priors Γ .
- ▶ Find $\min_{\Gamma} P(H_0 \mid x_{1:n})$: worst case for H_0 .
- ▶ Following Berger and Sellke (1987), we will consider:

$$\Gamma_A = \{\text{all distributions}\}$$

$$\Gamma_S = \{\text{all distributions symmetric about } \mu_0\}.$$

p -values are not $P(H_0 \mid x_{1:n})$

- ▶ Robust Bayes is *aggressive* against H_0 : we're picking the worst possible case over a class of priors.
- ▶ However, we'll see that

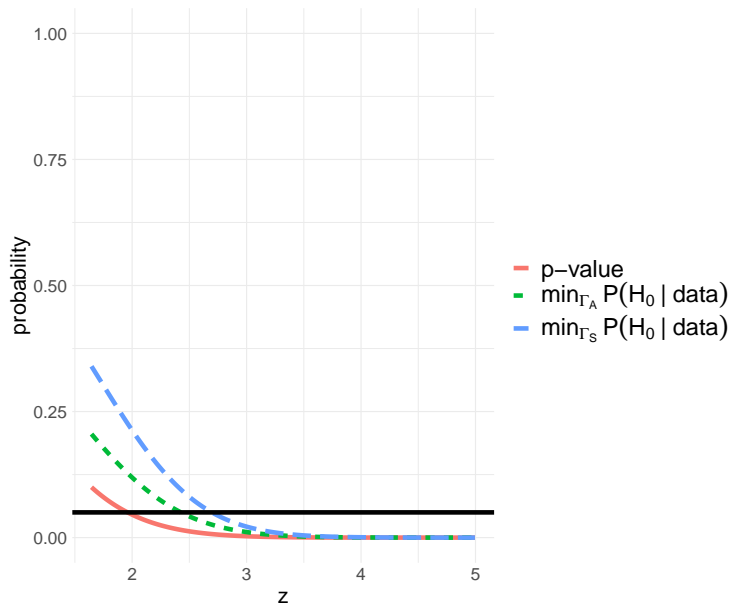
$$\min_{\Gamma} P(H_0 \mid x_{1:n}) \geq p\text{-value}.$$

- ▶ Interpreting a p -value as $P(H_0 \mid x_{1:n})$ overestimates evidence against H_0 .

Results: Table

z	p -value	$\min \Gamma_A$	$\min \Gamma_S$
1.645	0.100	0.205	0.340
1.960	0.050	0.128	0.227
2.576	0.010	0.035	0.068
3.291	0.001	0.004	0.009

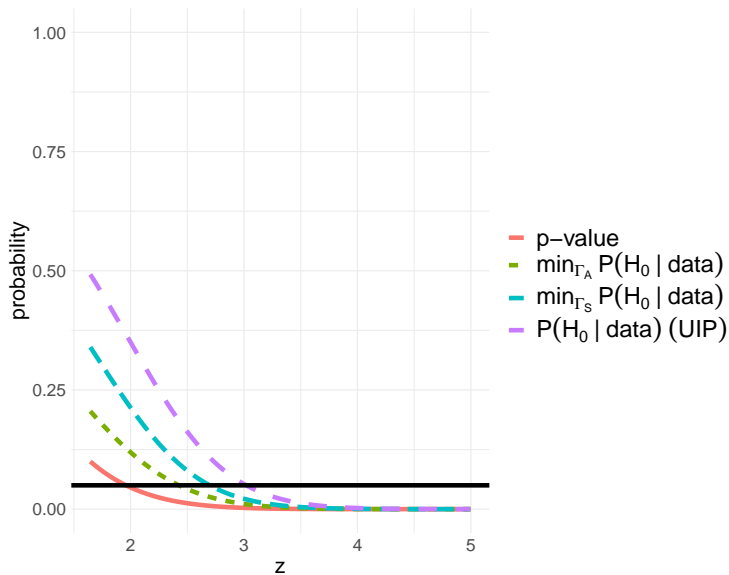
Results: Plot



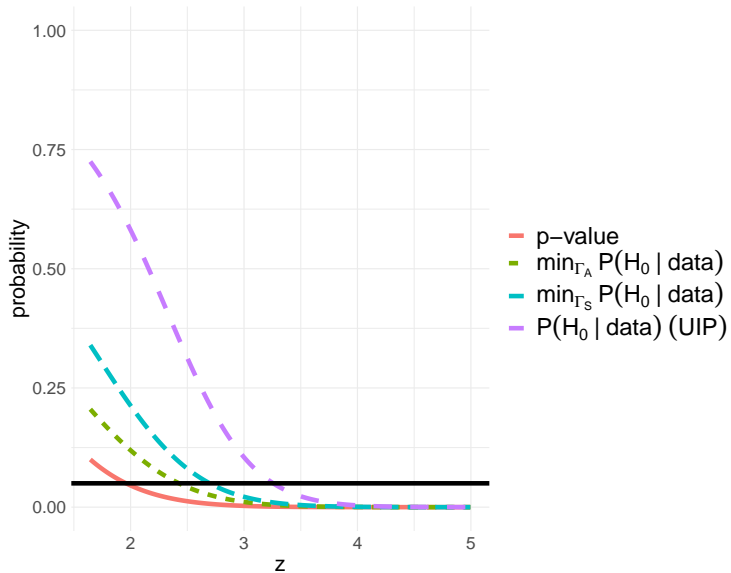
Comparison with unit information prior

- ▶ How far are robust Bayesian answers to what we would obtain with a default prior?
- ▶ Add in the results for the unit information prior.
- ▶ Results depend on z and n .
- ▶ The higher n is, the bigger the difference with p -values and robust Bayes.
- ▶ Similar results with other default priors: e.g., Cauchy or hyper- g .

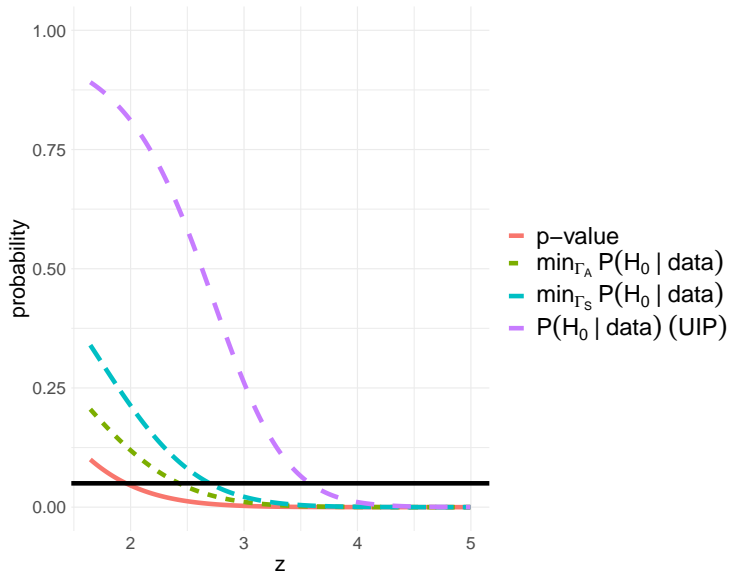
$n = 10$



$n = 100$



$n = 1000$



Conclusions

- ▶ $P(H_0 \mid x_{1:n})$ can be pretty different to p -values.
- ▶ In point null hypothesis testing of normal means,
$$P(H_0 \mid x_{1:n}) \geq p\text{-value}.$$

- ▶ Robust Bayes can be far from proper Bayes for large n .

Reference

Berger, J. O., & Sellke, T. (1987). Testing a point null hypothesis: The irreconcilability of p -values and evidence. *Journal of the American Statistical Association*, 82(397), 112-122.