BAYESIAN INCONSISTENCY UNDER MODEL MISSPECIFICATION

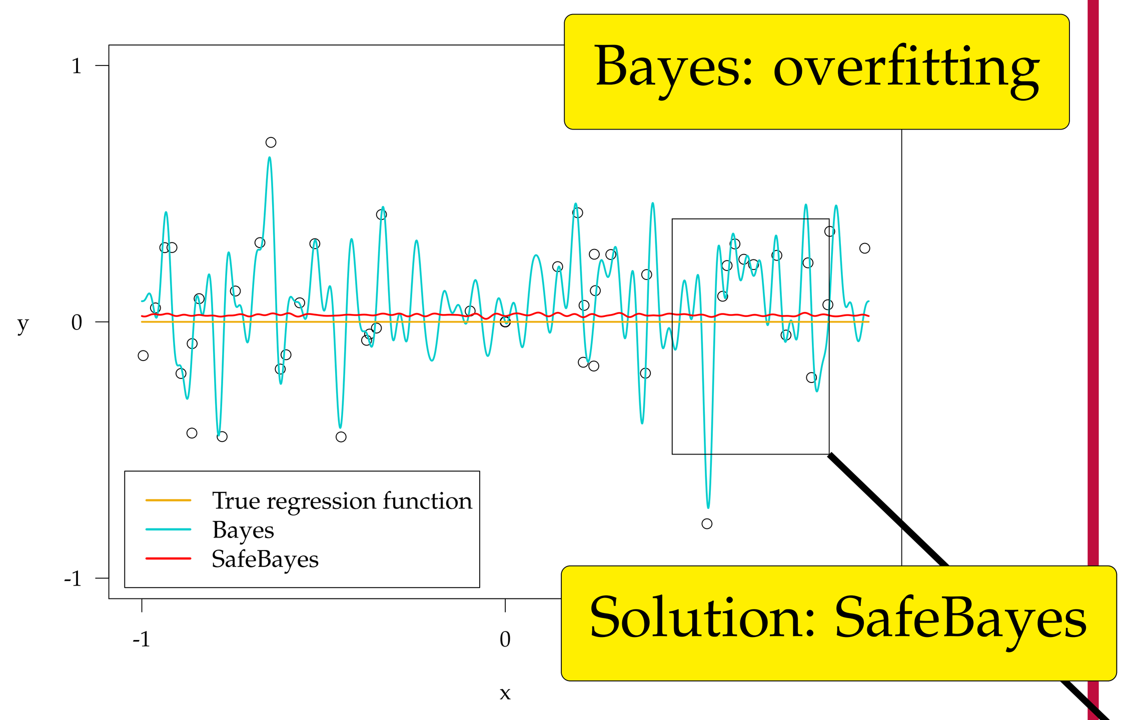
Centrum Wiskunde & Informatica

Rianne de Heide (heide@cwi.nl), joint work with Peter Grünwald and Thijs van Ommen

Universiteit Leiden

PROBLEM

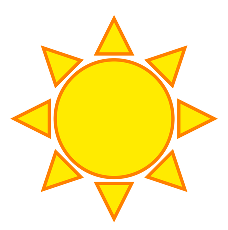
- Models are wrong but useful
- Extreme example: Bayesian inference can go very wrong when the model is misspecified
- Simple model: $y_i = \mathbf{f}(\mathbf{x_i}) + \epsilon_i$, $\epsilon_i \overset{iid}{\sim} \mathrm{N}(0, \frac{1}{4})$, Fourier basis
- Simple model misspecification: $y_i = \mathbf{0} + \epsilon_i$, $\epsilon_i \overset{iid}{\sim} \mathrm{N}(0, \frac{1}{4})$, $x_i \overset{iid}{\sim} \mathrm{U}(-1, 1)$, **but then** set half of the data to $(0, 0)$

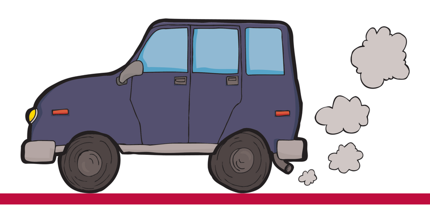


REAL WORLD DATA

Problem arises in real world data[1]:

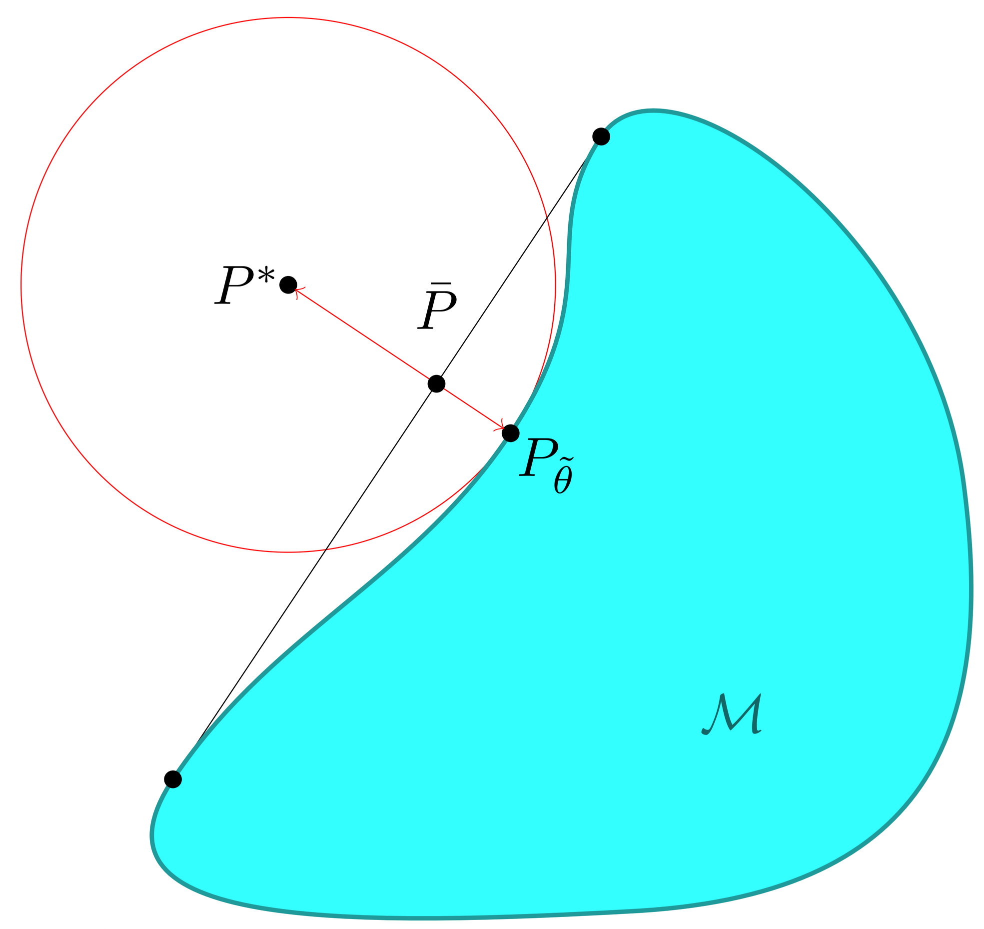
- Seattle weather data
- London air pollution





EXPLANATION

***Bad* misspecification**: $P_{\tilde{\theta}}$ is the closest distribution in the model $\mathcal{M}$ to the true distribution P^* in KL-divergence. Because the model is not convex, the Bayes predictive distribution $\overline{P}$ might be a mixture of *bad* distributions in the model that ends up outside $\mathcal{M}$.



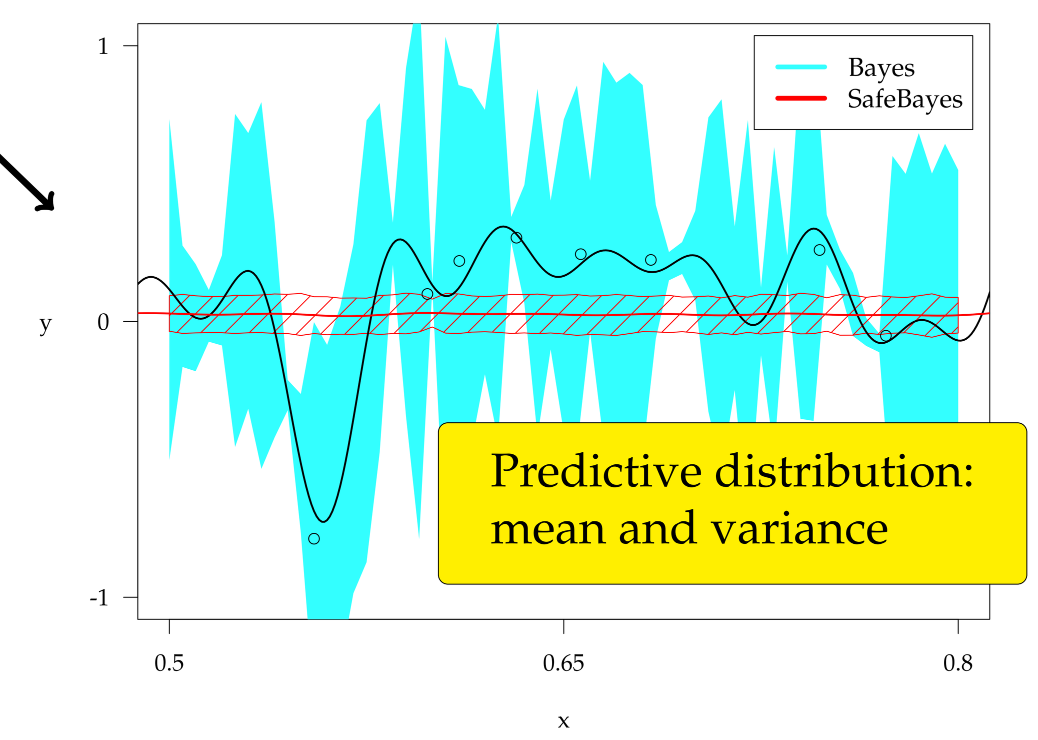
We have:

- Bad square-risk behaviour (experiments)
- Good log-risk behaviour from[4]:

$$\mathbf{E}_{Z^n \sim P^*}\left[\sum_{i=1}^{n}\left(\mathrm{RISK}^{\log}\,\overline{P}(\cdot\,|Z^{i-1}) - \mathrm{RISK}^{\log} P_{\tilde{\theta}}\right)\right] = \mathcal{O}(\log n)$$

This discrepancy implies that the posterior is not concentrated.

VARIANCE ISSUES



SOLUTION

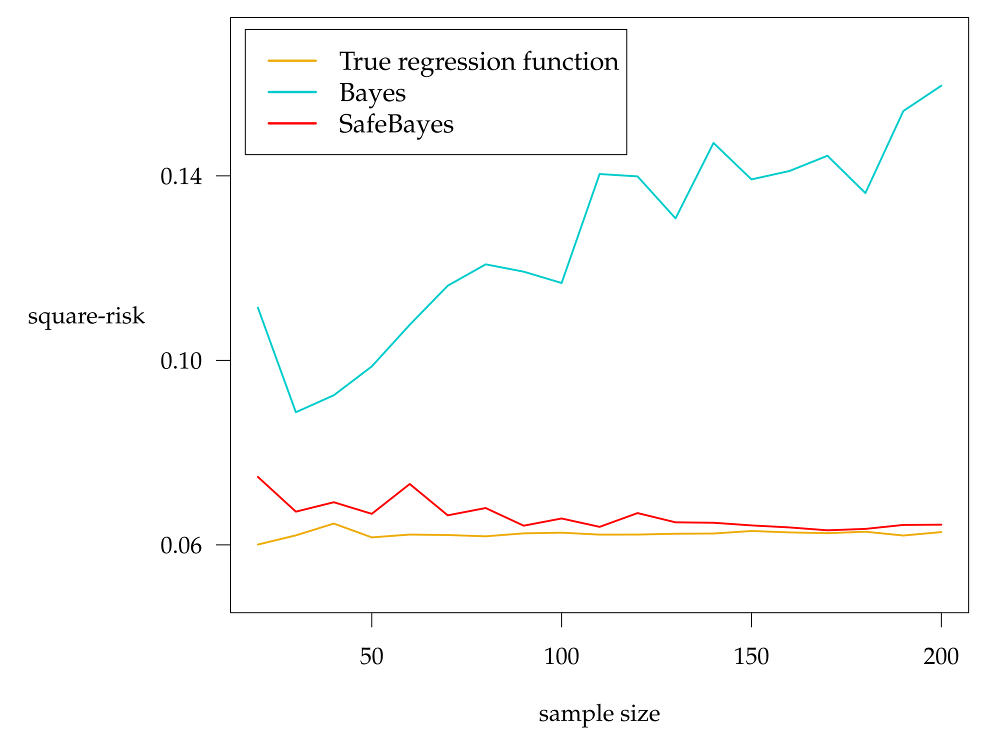
Generalized posterior

$$\pi(\theta|z^n, \eta) = \frac{(p(y^n|x^n, \theta))^\eta \pi(\theta)}{\int (p(y^n|x^n, \theta))^\eta \pi(\theta)\mu(d\theta)}$$

- deals with misspecification (posterior concentrates on $P_{\tilde{\theta}}$) if η taken *small enough*
- learn η with **The Safe-Bayesian algorithm**[3]
- good convergence rates[3], excellent performance in simulation setting and real world data[1,2, new work]

EXPERIMENTS

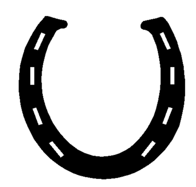
Bayes: Square-risk *increases* with sample size



SafeBayes performed in terms of square-risk
- never substantially worse
- sometimes substantially better

than *standard* Bayes in experiments with:
- Bayesian model averaging and selection
- Bayesian Ridge regression
- Bayesian Lasso regression
- Horseshoe regression

in different settings:
- multivariate
- polynomial
- Fourier basis



and with:
- different priors (Jeffreys', Raftery's, slightly informative priors, etc.)
- fixed and varying variance
- variations on the simulated data (less *easy points*, less noise, etc.)
- real world data

REFERENCES

[1] R. de Heide. The Safe-Bayesian Lasso, Master Thesis, University of Leiden (2016)

[2] P.D. Grünwald, T. van Ommen. Inconsistency of Bayesian Inference for Misspecified Linear Models, and a Proposal for Repairing It, arXiv:1412.3730 (2014)

[3] P.D. Grünwald. The Safe Bayesian, In ALT Proceedings, Pp. 169-183. (2012)

[4] A.R. Barron. Information-theoretic characterization of Bayes performance and the choice of priors in parametric and nonparametric problems, In Bayesian Statistics, Vol 6, Pp. 27-52. (1998)

SOFTWARE

`SafeBayes`

`R` package available on `CRAN`.