Bayesian analysis of homogeneity studies in the production of reference materials

Adriaan M.H. van der Veen

Van Swinden Laboratorium, Delft, the Netherlands

21 April 2018
CITAC Best Papers Award 2017, Sévres, France

Rationale

- Well known that the repeatability standard deviation (σ) sets a lower limit to the determination of the between-bottle standard deviation (τ)
- Increase of the number of replicates not always possible
- Need for a statistically sound approach for monitoring the performance of a decanting method for synthetic natural gas mixtures

Repeatability standard deviations of the natural gas analyser

Table: Amount-of-substance fractions (x) expressed in % and repeatability relative standard deviations (s) for natural gas

	x	s
ethane	3.4978	0.012 %
propane	4.7007	0.010 %
iso-butane	0.1234	0.011 %
n-butane	0.1225	0.012 %
iso-pentane	0.0306	0.031 %
n-pentane	0.0312	0.021 %
n-hexane	0.0760	0.014 %
carbon dioxide	5.0327	0.023 %
nitrogen	0.4251	0.165 %
methane	85.9601	0.011 %

Computation of the between-bottle standard deviation

- Compute the between-group variance

$$\tau^2 = \frac{MS_{\text{between}} - MS_{\text{within}}}{n}$$

- Problems arise when $MS_{\text{between}} < MS_{\text{within}}$

Table: ANOVA results for nitrogen

Source of Variation	*SS*	ν	*MS*	*F*
Between Groups	2.92×10^{-10}	9	3.25×10^{-11}	0.663
Within Groups	1.96×10^{-9}	40	4.90×10^{-11}	
Total	2.25×10^{-9}	49		

Alternatives if $MS_{\text{between}} < MS_{\text{within}}$

- set $\tau = 0$ (e.g., ISO 13528:2015, ISO Guide 35:2017)
- set $\tau = \sqrt{\frac{MS_{\text{within}}}{n}}$ (Ellison, 2015)
- set $\tau = \sqrt{\frac{MS_{\text{within}}}{n}} \sqrt[4]{\frac{2}{\alpha(n-1)}}$
  (Van der Veen et al., 2001; ISO Guide 35:2017)

Bayesian modelling

- “Solution” published in 2001 has never been viewed as entirely satisfactory
- None of the three options permits a judgment on the performance of the decanting method
- Bayesian inference will provide a posterior probability distribution for τ
- Mean of the probability distribution of τ can be taken as the estimate for the between-bottle standard deviation

Bayesian model (1)

- Same underlying assumptions for the data
  - normally distributed data
  - homogeneity of the within-group standard deviations
  - no substantial prior information
- Same statistical model as used in 2001 with classical analysis of variance

$$y_{ij} = \mu + B_i + \varepsilon_{ij}$$

where y_{ij} denotes the jth replicate on the ith mixture, μ the amount-of-substance fraction, B_i the bias of ith mixture, and ε_{ij} the measurement error

Bayesian model (2)

- Model parameters:
  - Batch value for the amount-of-substance fraction μ
  - Within-group variance $\sigma^2 = \mathrm{var}(\varepsilon_{ij})$
  - Between-group variance $\tau^2 = \mathrm{var}(B_i)$
- For each of the parameters, a prior probability distribution is to be assigned

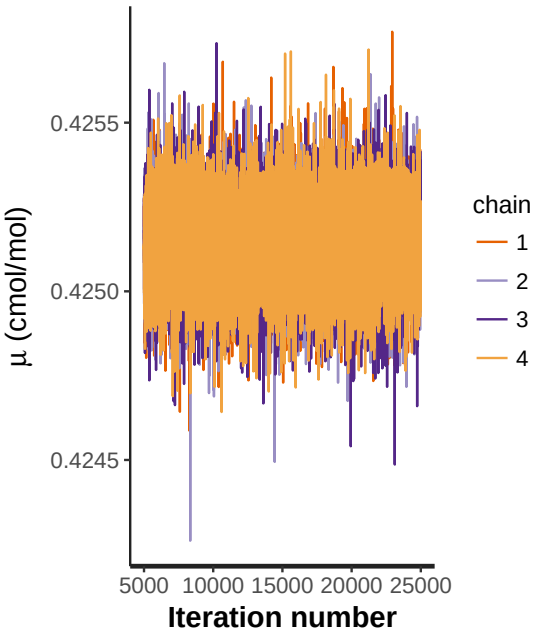
Prior distributions

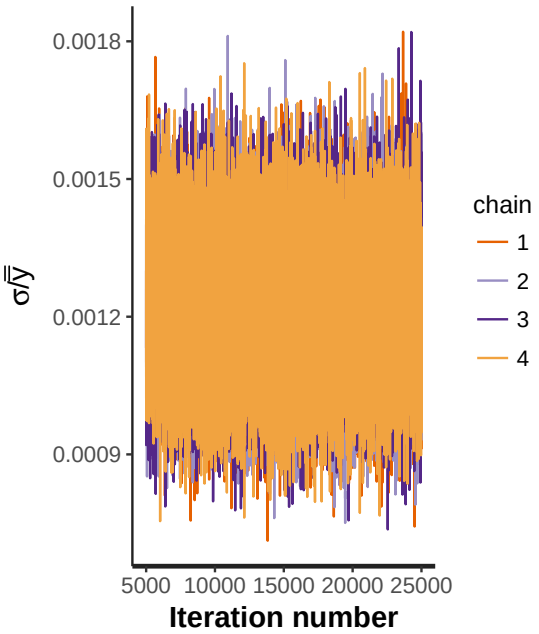
- Non-informative priors do not really reflect the information available
- Weakly informative priors
  - allow a (weak) influence of our prior information (nominal amount-of-substance fraction, repeatability standard deviation, between-bottle standard deviation)
  - improve performance of the Markov Chain Monte Carlo method (MCMC)
  - do not require proving that the posterior probability distribution is “proper”
  - should be assessed for actually behaving as “weakly informative”

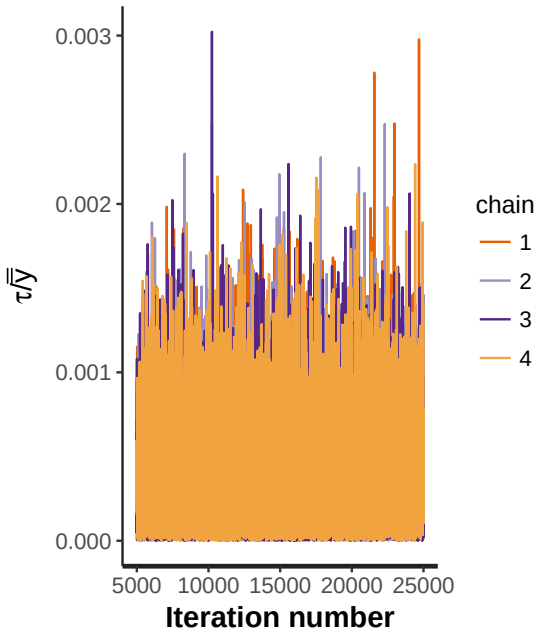
Bayesian model (3)

- $\mu \sim \mathrm{N}(\mu_0, \sigma_0^2)$; $\sigma_0 = 0.05\mu_0$
- $\sigma \sim \mathrm{Cauchy}(0, A_\sigma)$; $A_\sigma = 2\max_i(s_i)$
- $\tau \sim \mathrm{Cauchy}(0, A_\tau)$; $A_\tau = 5s_1$
- Parameters for the folded Cauchy-distribution (t-distribution with $\nu = 1$) set sufficiently wide to ensure "weakly informative" behaviour

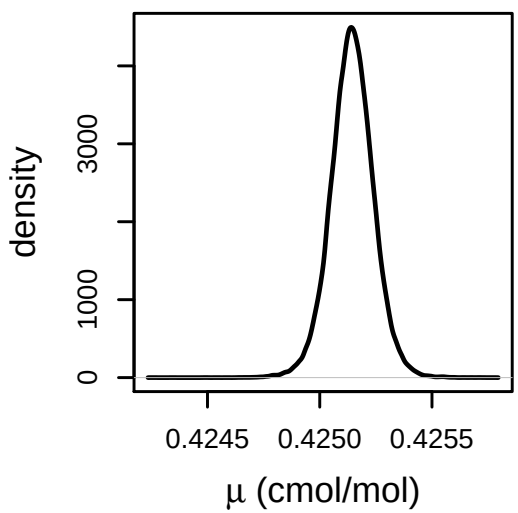
Trace plots nitrogen

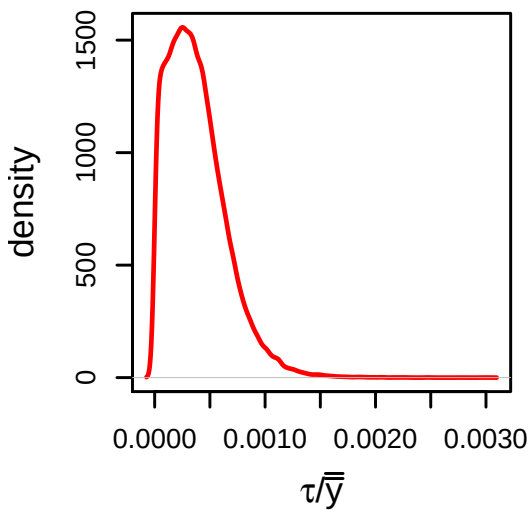


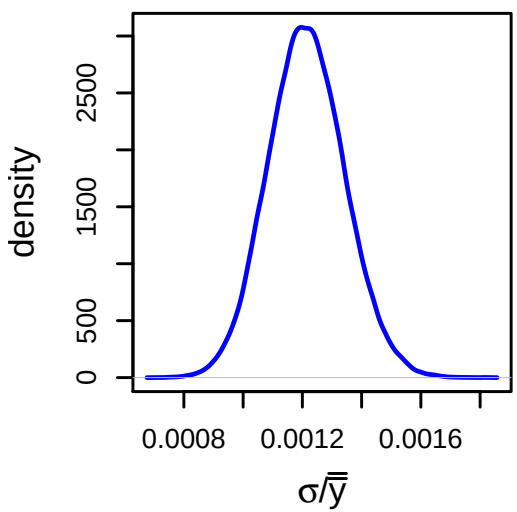




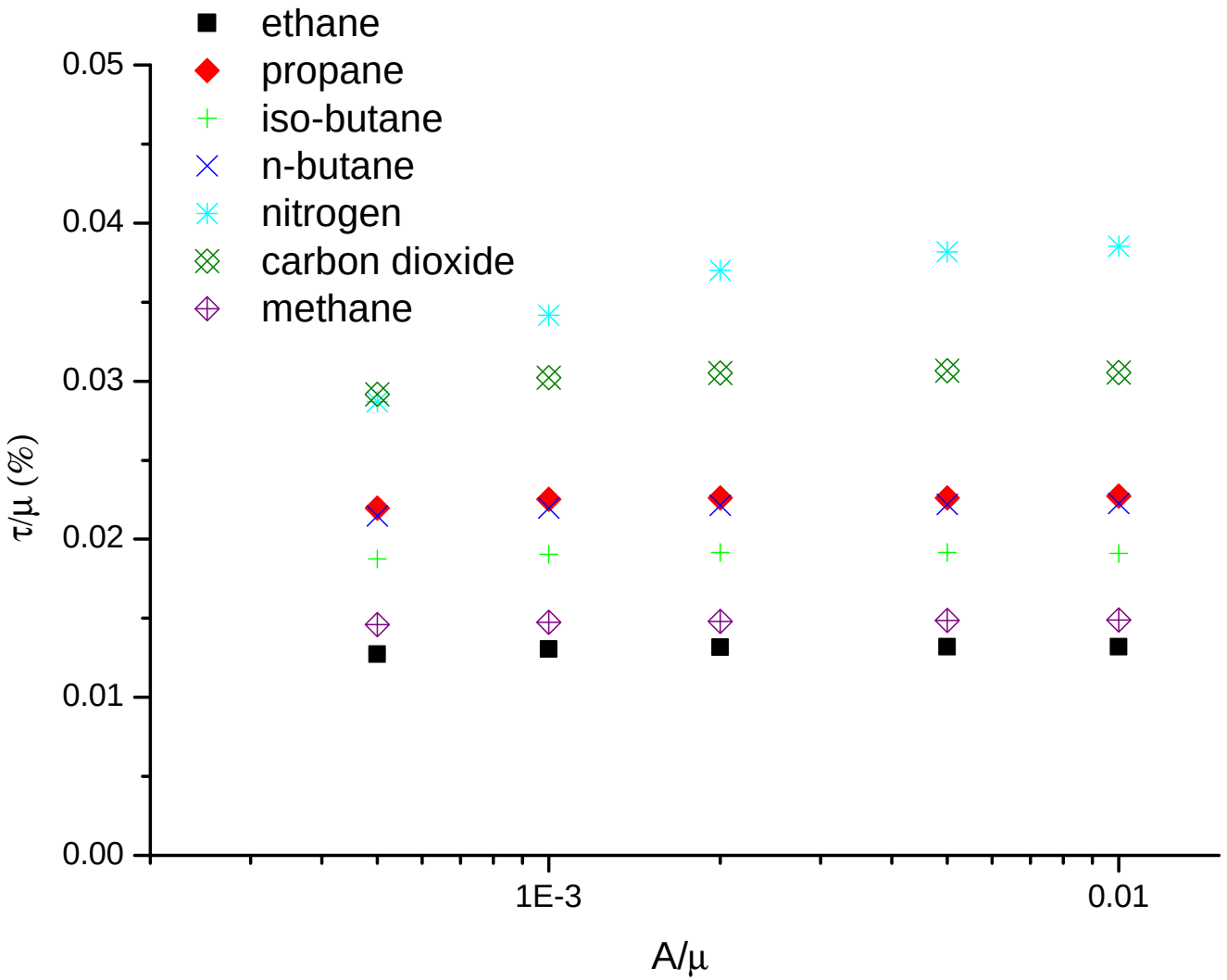
Posteriors for μ, σ and τ for nitrogen







Sensitivity analysis



Nitrogen: results and the alternatives

Table: Results of the MCMC calculation for nitrogen.

	mean	sd	$q_{0.025}$	$q_{0.975}$
μ	0.425 144 3	0.000 168 0	0.424 812 2	0.425 475 7
τ	0.000 172 0	0.000 143 9	0.000 006 1	0.000 532 0
σ	0.000 480 8	0.000 055 7	0.000 375 2	0.000 593 5

- Bayesian model: $\hat{\tau} = 1.72\,\mu\text{mol}\,\text{mol}^{-1}$, which lies in between
- ISO Guide 35: $u^*_{\text{bb}} = 1.48\,\mu\text{mol}\,\text{mol}^{-1}$ and
- Elisson: $\sqrt{MS_{\text{within}}/n} = 3.13\,\mu\text{mol}\,\text{mol}^{-1}$.
- Both u^*_{bb} and $\sqrt{MS_{\text{within}}/n}$ within the 95 % coverage interval of τ.

Conclusions

- Bayesian analysis is to be preferred where classical analysis of variance and meta-analysis fail
- Computations are involved (require Markov Chain Monte Carlo method)
- Bayesian inference provides a wealth of information, well beyond GUM requirements
- Classical methods compare well with the Bayesian model proposed, as long as $MS_{\text{between}} > MS_{\text{within}}$