Universiteit Utrecht

A Gentle Introduction to Bayesian Statistics

Rens van de Schoot
Sara van Erp
Duco Veen
Beth Grandfield

Learning goal:

Overview

Day 1 : Conceptual introduction

Day 2 : WAMBS-checklist (when to worry and how to avoid the misuse of Bayesian Statistics)

Day 3 : Estimation methods including alternatives that can be more efficient when dealing with computational or non-covergence issues (MCMC, Gibbs, MH, HMC, NUTS, etc.)

Day 4 : Prior sensitivity analysis to investigate the influence the prior has on the results; models with many parameters; shrinkage priors.

Day 5 : Informative priors; expert knowledge.
We end with general reflections.

Time Schedule Day 1-5:

0900-1200: lecture
1200-1300: lunch
1330-1500: (supervised) computer lab
1500-1600: Q&A

Software:

Table 2 A non-exhaustive summary of commonly used and open Bayesian software programs

From: Bayesian statistics and modelling

Software package	Summary
General-purpose Bayesian inference software	
BUGS[231,232]	The original general-purpose Bayesian inference engine, in different incarnations. These use Gibbs and Metropolis sampling. Windows-based software (WinBUGS[233]) with a user-specified model and a black-box MCMC algorithm. Developments include an open-source version (OpenBUGS[234]) also available on Linux and Mac
JAGS[235]	An open-source variation of BUGS that can run cross-platform and can run from R via rjags[236]
PyMC3[237]	An open-source framework for Bayesian modelling and inference entirely within Python; includes Gibbs sampling and Hamiltonian Monte Carlo
Stan[98]	An open-source, general-purpose Bayesian inference engine using Hamiltonian Monte Carlo; can be run from R, Python, Julia, MATLAB and Stata
NIMBLE[238]	Generalization of the BUGS language in R; includes sequential Monte Carlo as well as MCMC. Open-source R package using BUGS/JAGS-model language to develop a model; different algorithms for model fitting including MCMC and sequential Monte Carlo approaches. Includes the ability to write novel algorithms
Programming languages that can be used for Bayesian inference	
TensorFlow Probability[239,240]	A Python library for probabilistic modelling built on Tensorflow[203] from Google
Pyro[241]	A probabilistic programming language built on Python and PyTorch[204]
Julia[242]	A general-purpose language for mathematical computation. In addition to Stan, numerous other probabilistic programming libraries are available for the Julia programming language, including Turing.jl[243] and Mamba.jl[244]
Specialized software doing Bayesian inference for particular classes of models	
JASP[245]	A user-friendly, higher-level interface offering Bayesian analysis. Open source and relies on a collection of open-source R packages
R-INLA[230]	An open-source R package for implementing INLA[246]. Fast inference in R for a certain set of hierarchical models using nested Laplace approximations
GPstuff[247]	Fast approximate Bayesian inference for Gaussian processes using expectation propagation; runs in MATLAB, Octave and R

Software:

Day 1&5:
- Online apps

Day 2,3,4:
- R (brms)

"... it is clear that it is not possible to think about learning from experience and acting on it without coming to terms with Bayes' theorem."

Jerome Cornfield (in de Finetti, 1974a*)*

"**…**whereas the 20th century was dominated by NHST [null hypothesis significance testing], the 21st century is becoming Bayesian**…**"

Kruschke (2011, p.272) in a special 'Bayesian' issue of *Perspectives on Psychological Science*

"..."[...] seven decades of criticism against NHST is finally having an effect.

Sohlberg & Andersson (2005, p.69)

"…"[...] seven decades of criticism against NHST is finally having an effect.

"Besides correcting the most obvious flaws of NHST in a manner reminiscent of how meta-analysis does it, Bayesian statistics can be seen as providing answers to the questions researchers would be asking, unless they had first been taught a flawed method**…"**

Sohlberg & Andersson (2005, p.69)

"…Over the last few decades, it has become the major approach in the field of statistics, and has come to be accepted in many or most of the physical, biological and human sciences…"

Lee (2011, p.1)

It all started…

In 1748 when Hume published an essay about uncertainty

This essay inspired Thomas Bayes (1701-1761) who was enrolled at the University of Edinburgh to study logic and theology

He worked on the question whether God exists using Inverse Probability, but he never published any work on this topic

After T. Bayes passed, his relatives asked Richard Price (1723-1791) to go through his unfinished work and it was Price who discovered the paper on inverse probability

LII. *An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr.* Bayes, *communicated by Mr.* Price, *in a letter to* John Canton, *M. A. and F. R. S.*

Dear Sir,

Read Dec. 23, 1763. I now send you an essay which I have found among the papers of our deceased frie
and well deserves to
nearly interested in
particular reason for
cannot be improper.

He had, you kn
ciety, and was much esteemed by many as a very able mathema
introduction which he has writ to this Essay, he says, that his des … rst in thinking on the subject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know nothing concerning it but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times. He adds, that he soon perceived that it would not be

This is about **inverse probability**:
assigning a probability distribution to an unobserved variable.

with many further successes, and that you may enjoy every valuable blessing, is the sincere wish of, Sir,

your very humble servant,
Richard Price.

Newington Green,
Nov. 10, 1763.

Pierre Simon Laplace (1749-1827)

Independently discovered the same theorem and actually published the formula we now know as Bayes' rule…

(he also published the central limit theorem)

Bayes goes to war….

Used for artillery testing during Napoleon war

Bayes goes to war....

Testing ammunition during WOI

Frequentist methods required too much losses

Bayes goes to war….

Alan Turing

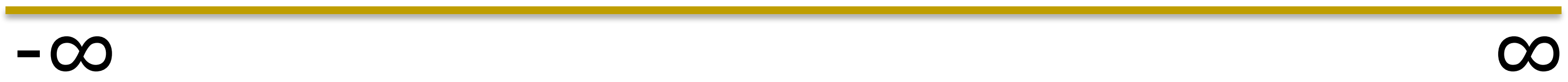

Prior Knowledge
Data


Prior Knowledge
Data

Bayes' rule:

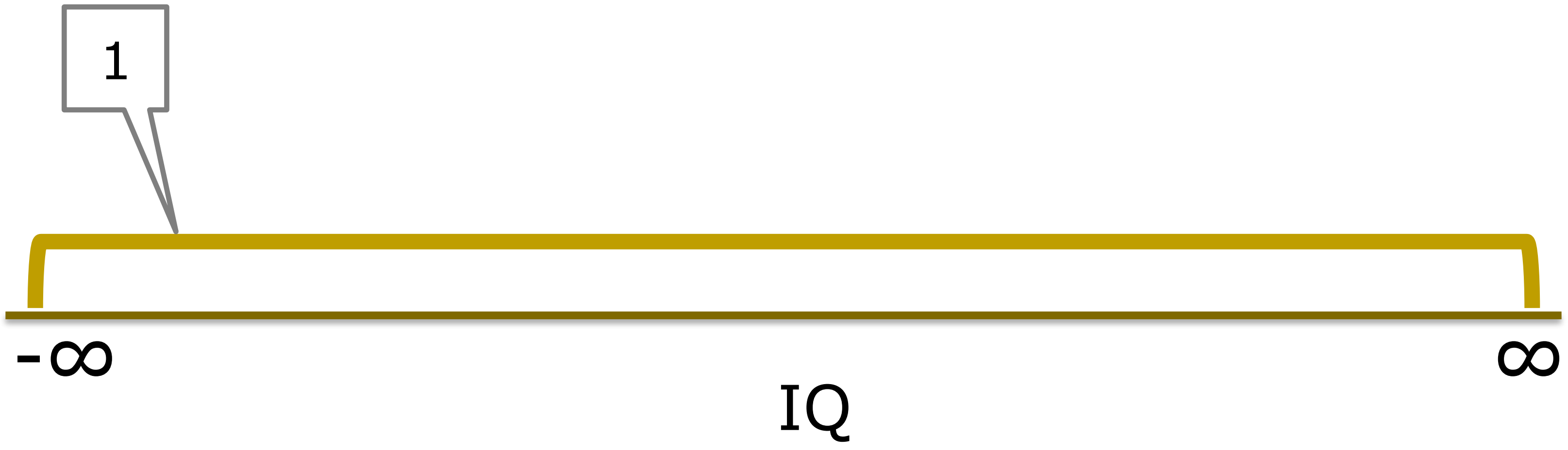
$$P(H \mid e) = \frac{P(e \mid H)\, P(H)}{P(e)}$$

Picture taken from:
http://www.psychologyinaction.org/2012/10/22/bayes-rule-and-bomb-threats/

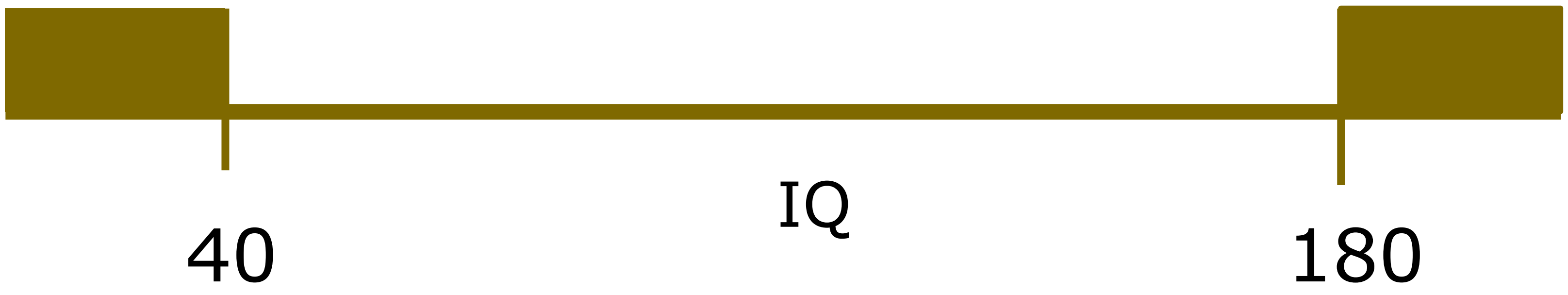


A Gentle Introduction to Bayesian Analysis: Applications to Developmental Research (2014). Van de Schoot, Kaplan, Denissen, Asendorpf, Neyer, van Aken. *Child Development*

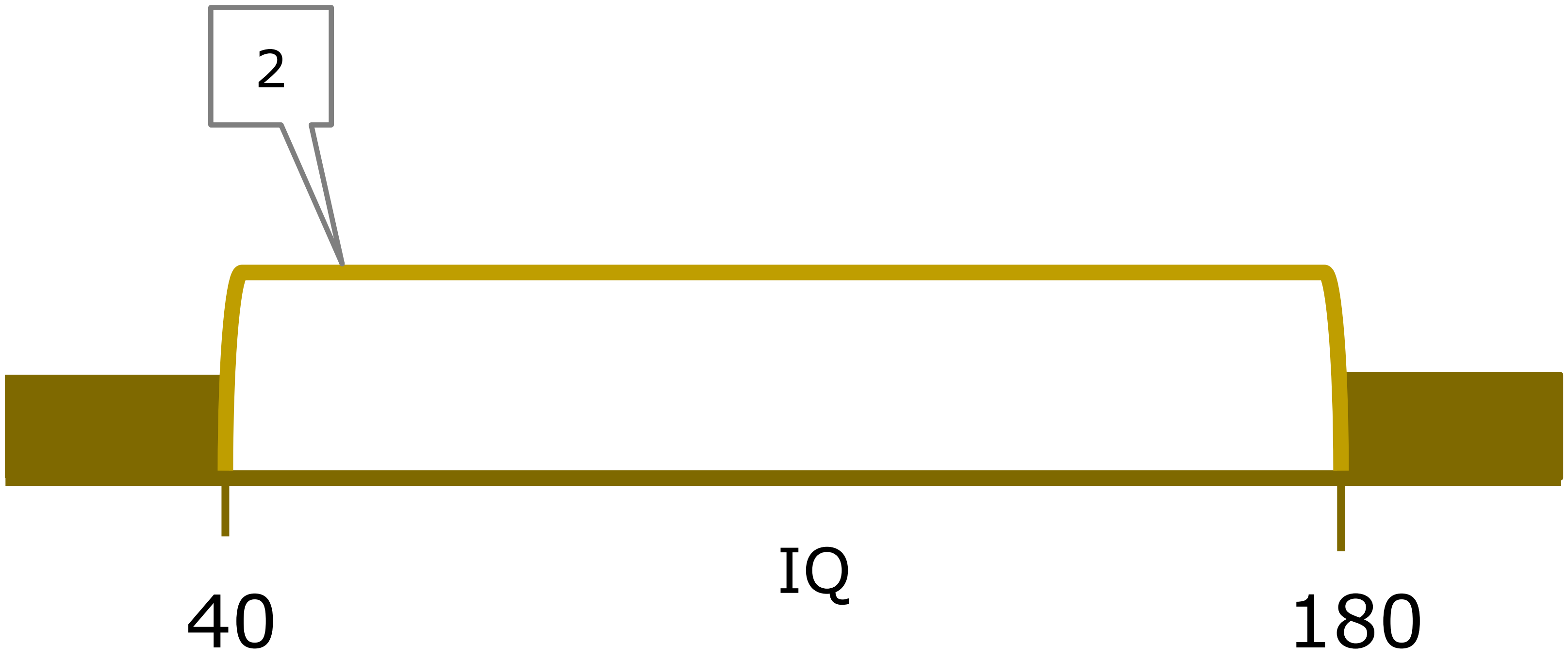
Prior Knowledge



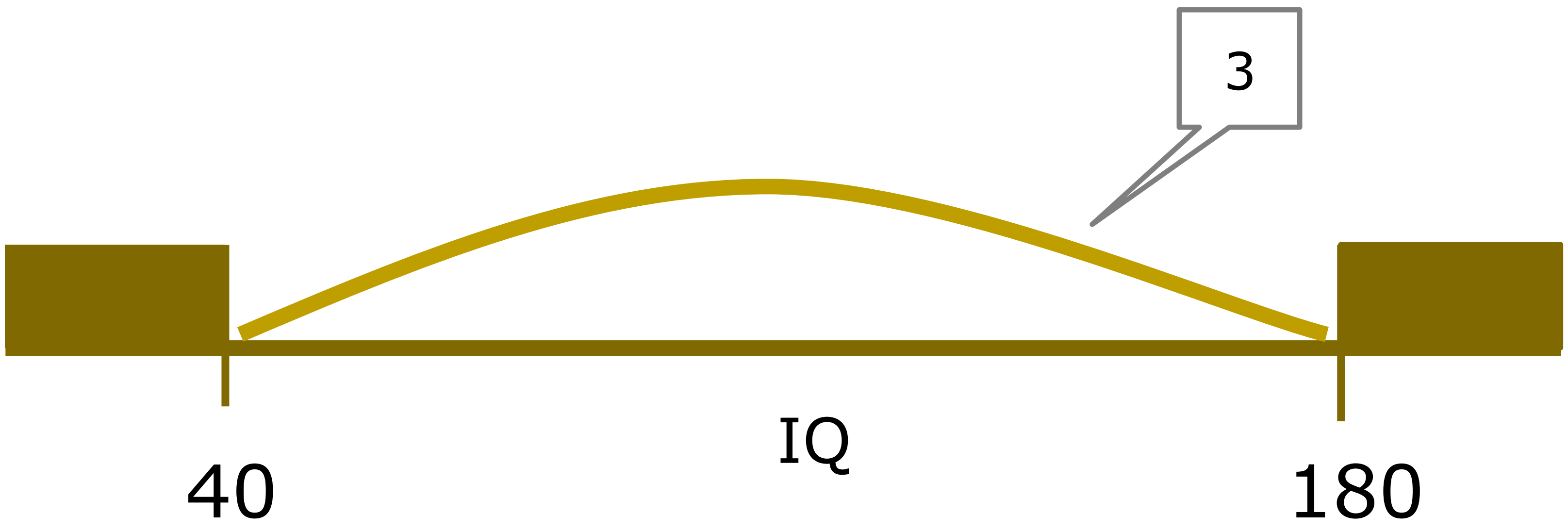
Prior Knowledge



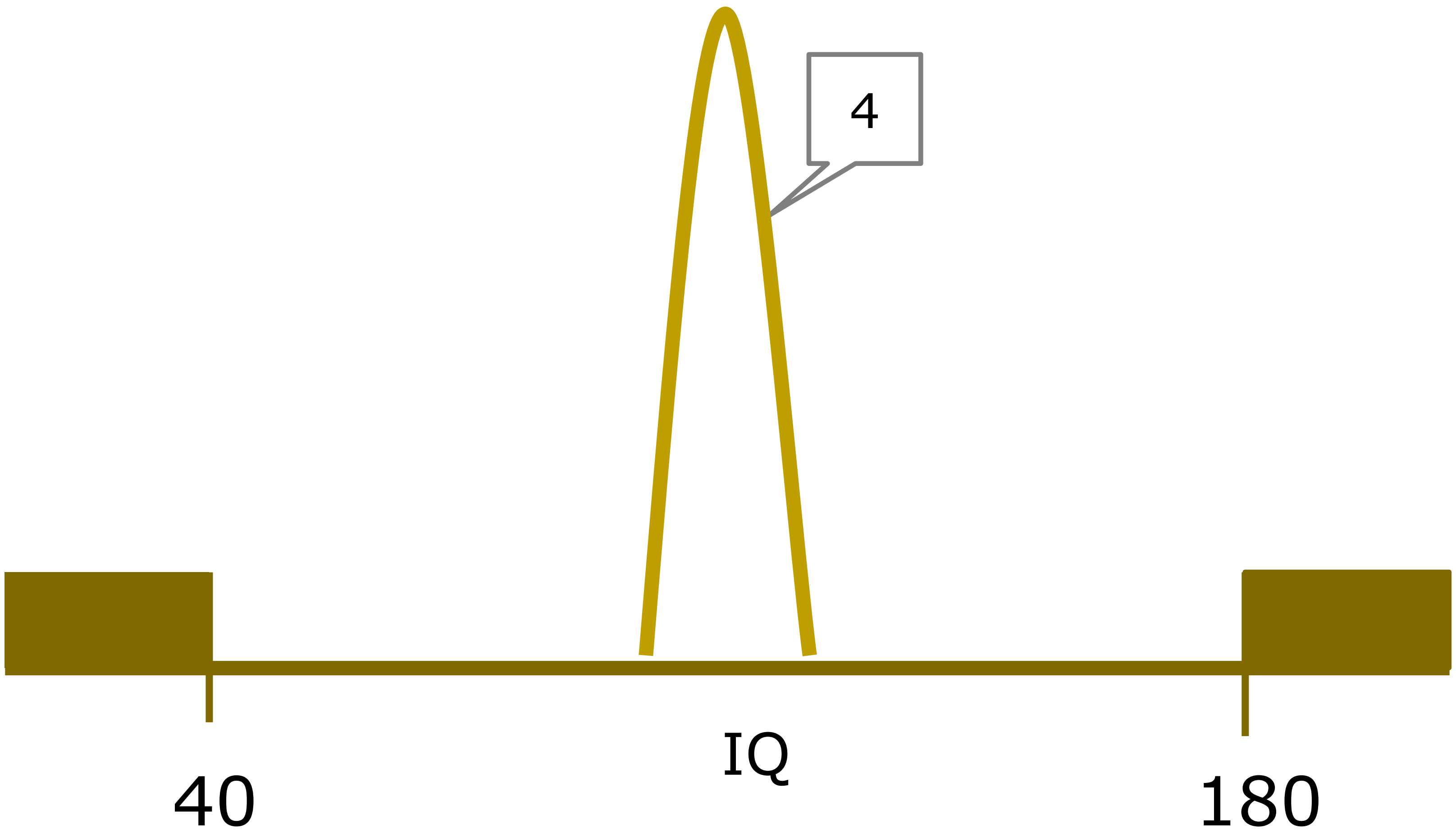
Prior Knowledge



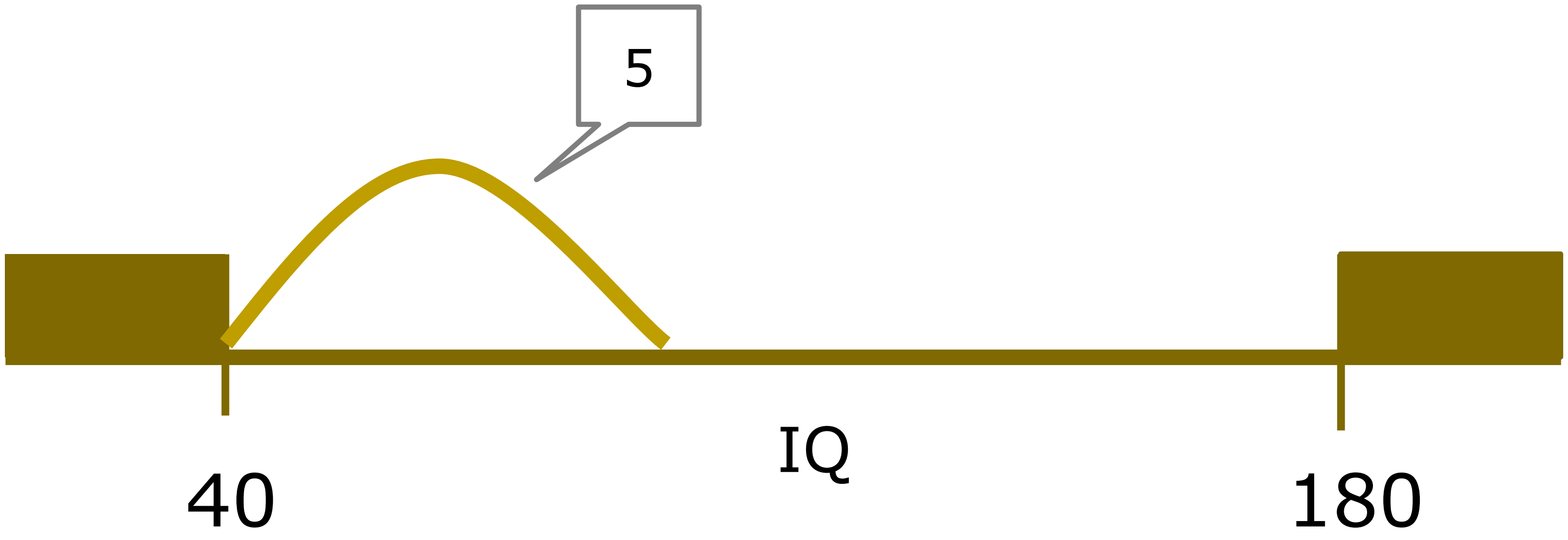
Prior Knowledge



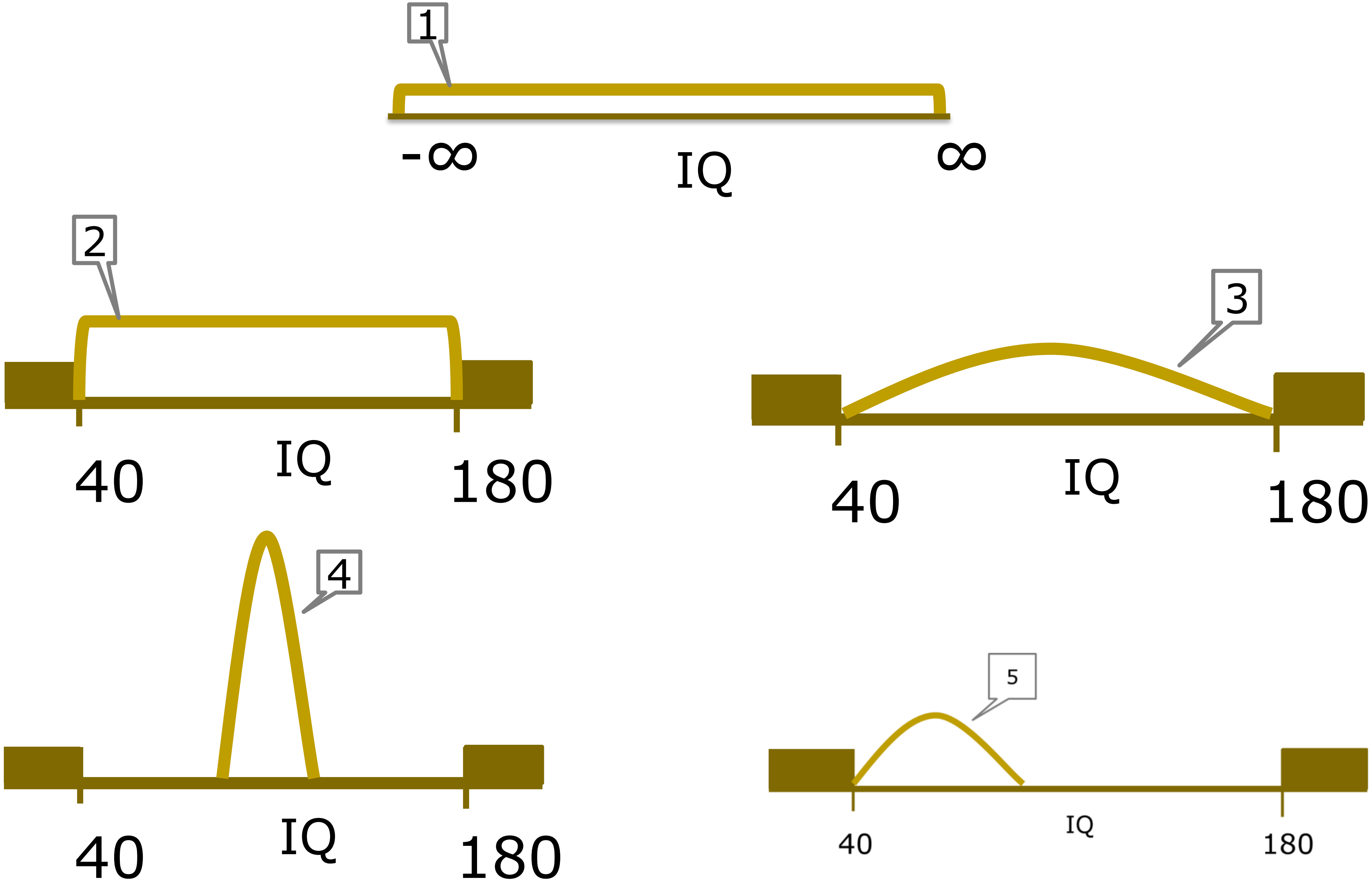
Prior Knowledge

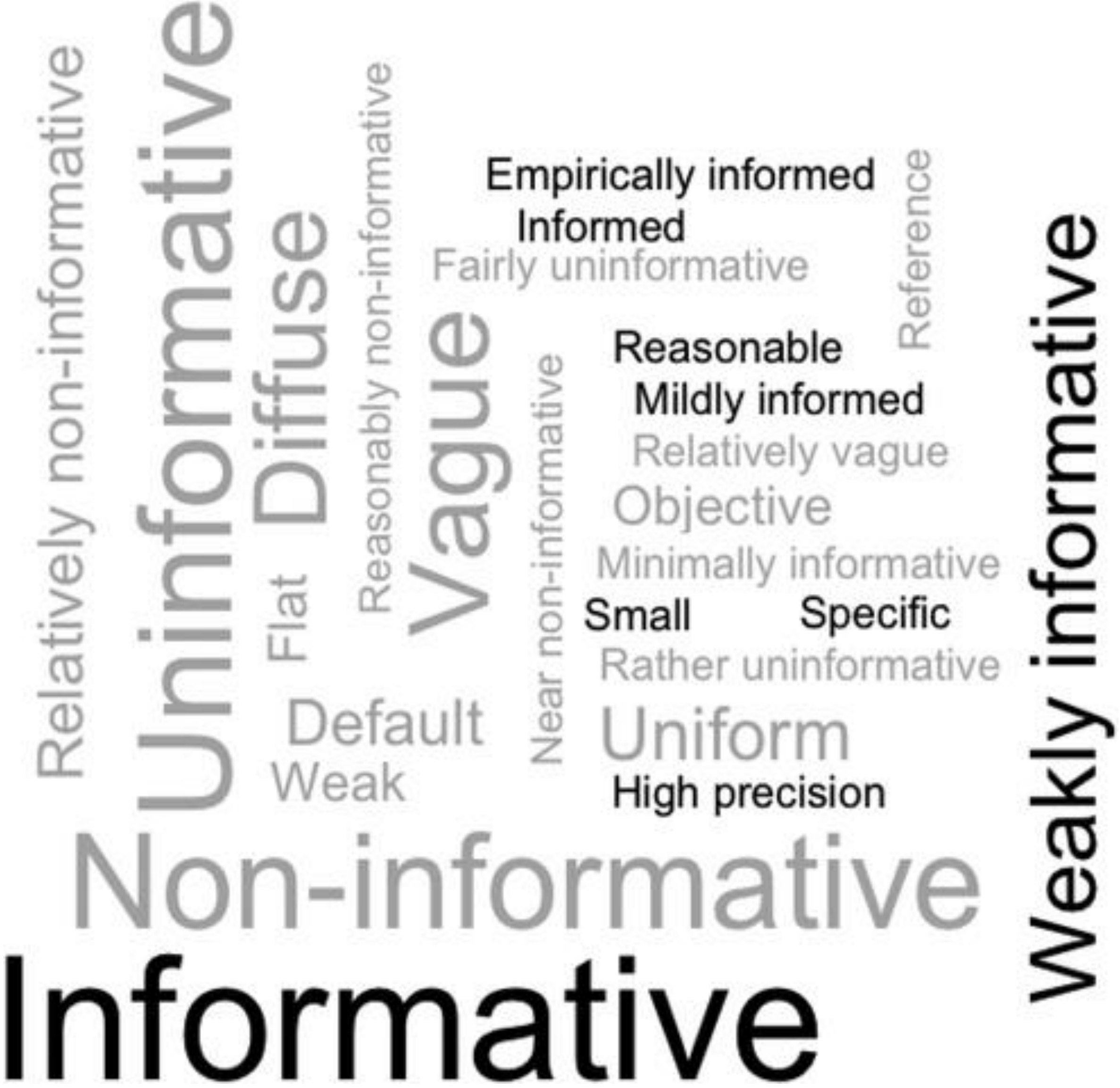


Prior Knowledge



Prior Knowledge





Wordcloud showing terms used to describe the level of informativeness of the priors in the empirical regression-based articles.

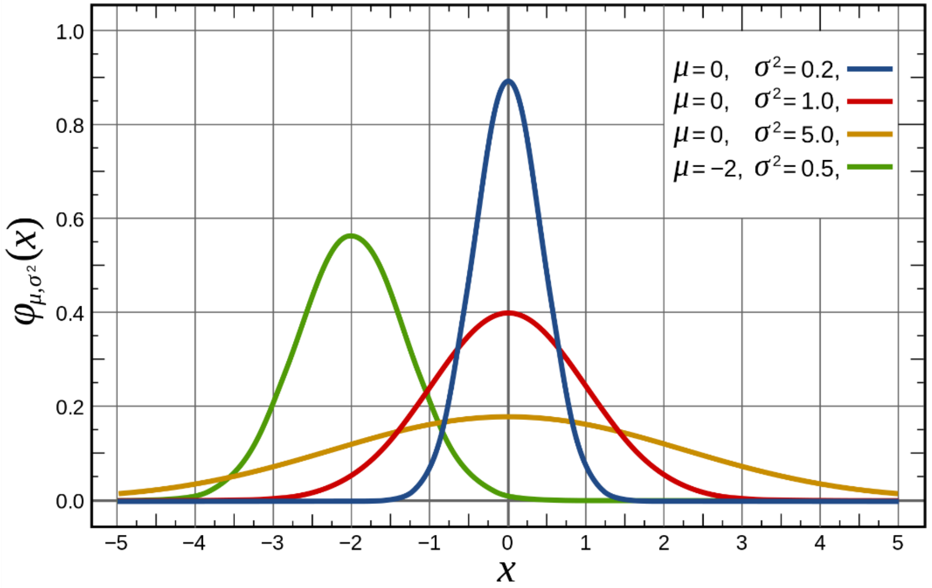
van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods, 22*(2), 217-239. http://dx.doi.org/10.1037/met0000100

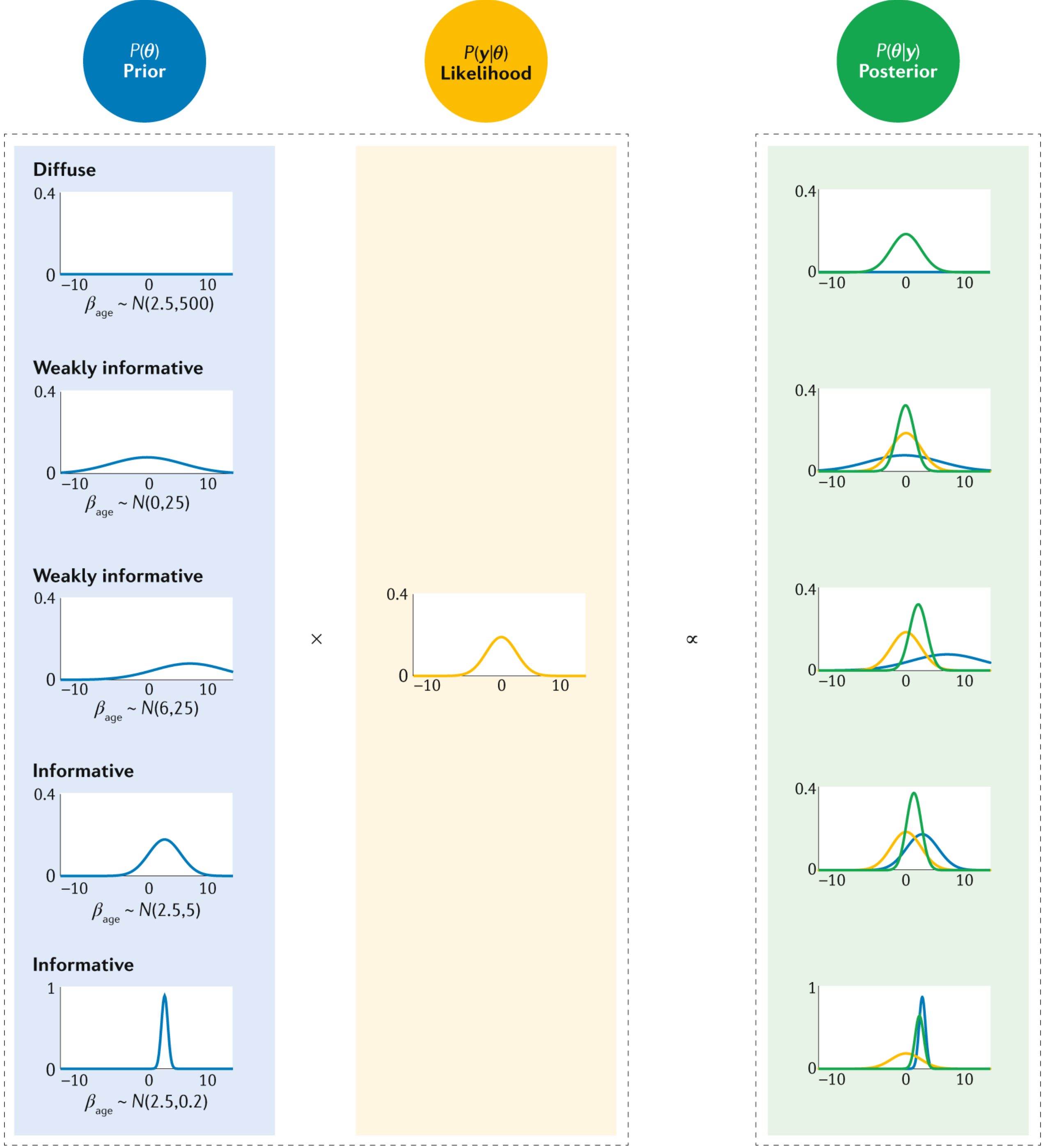
Choosing a prior

Step 1: Type of prior

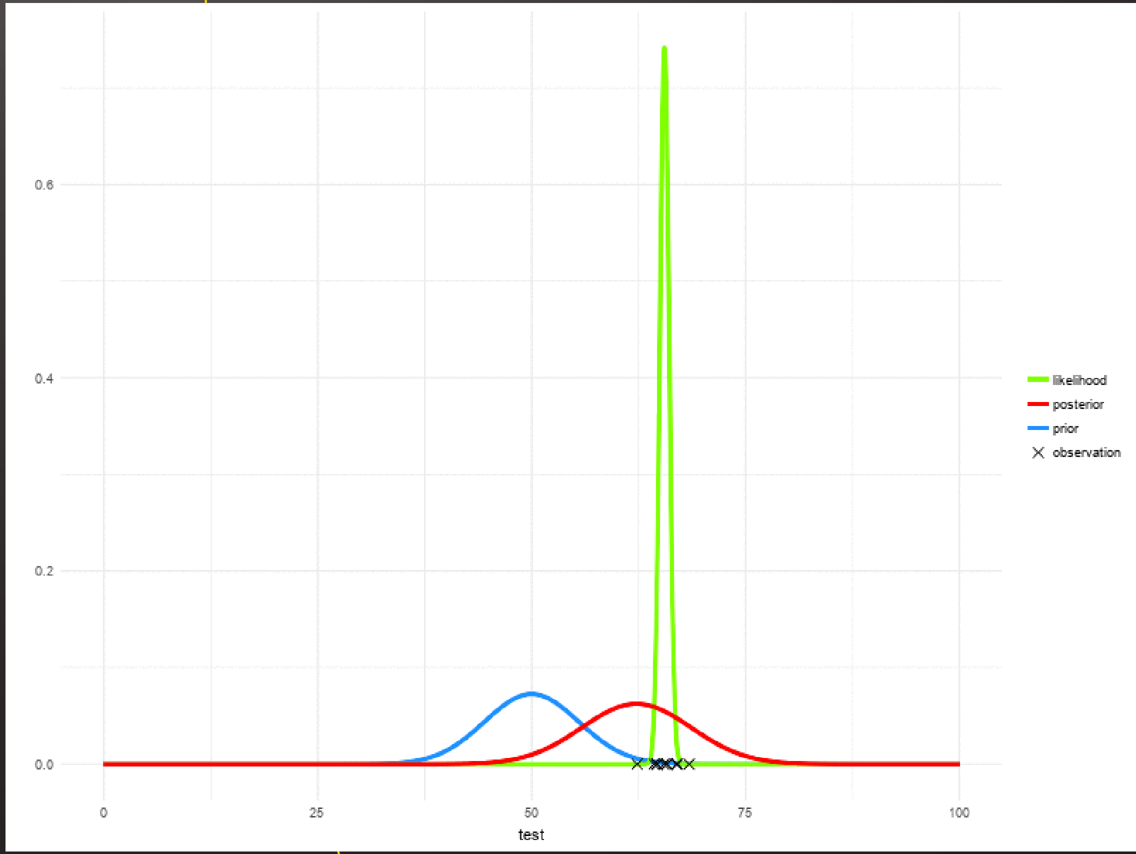
normal, gamma, chi2, wishart, binominal, Jeffreys'prior, uniform, beta, Laplace prior, AND MANY MANY MORE

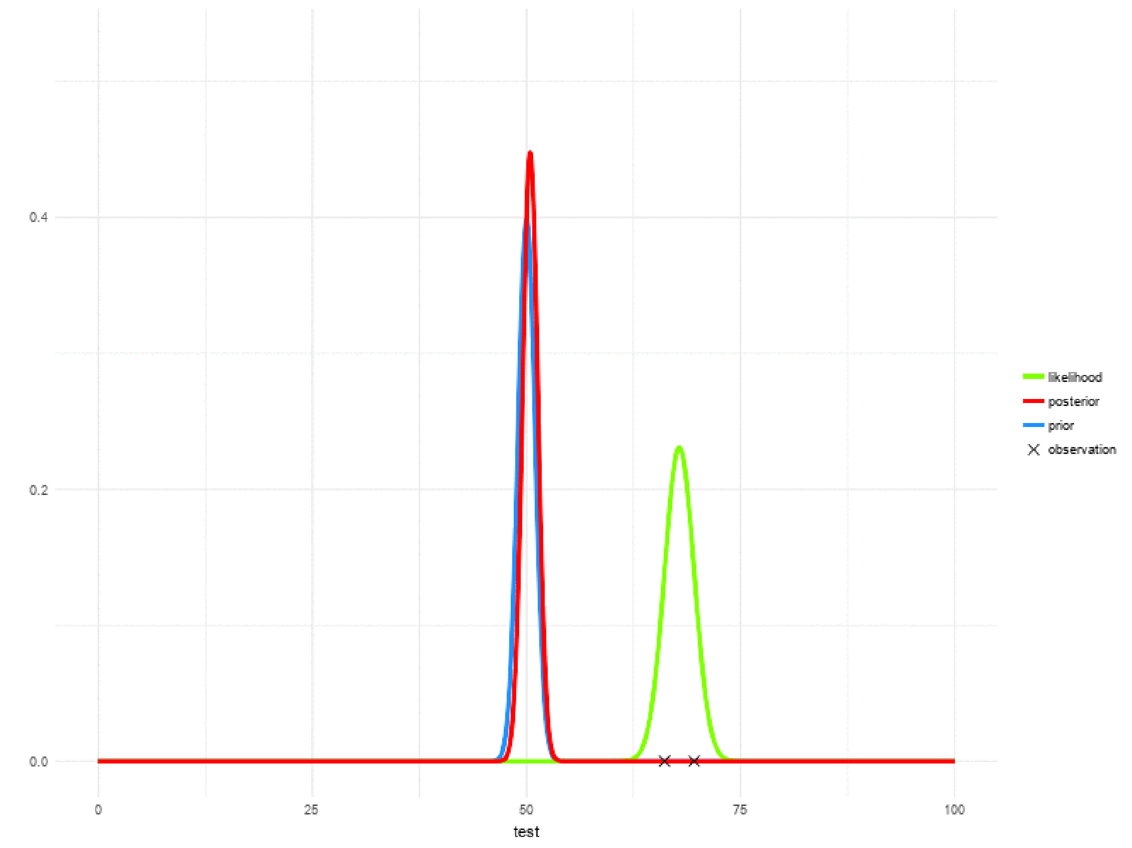
Step 2: Specify the hyper parameters





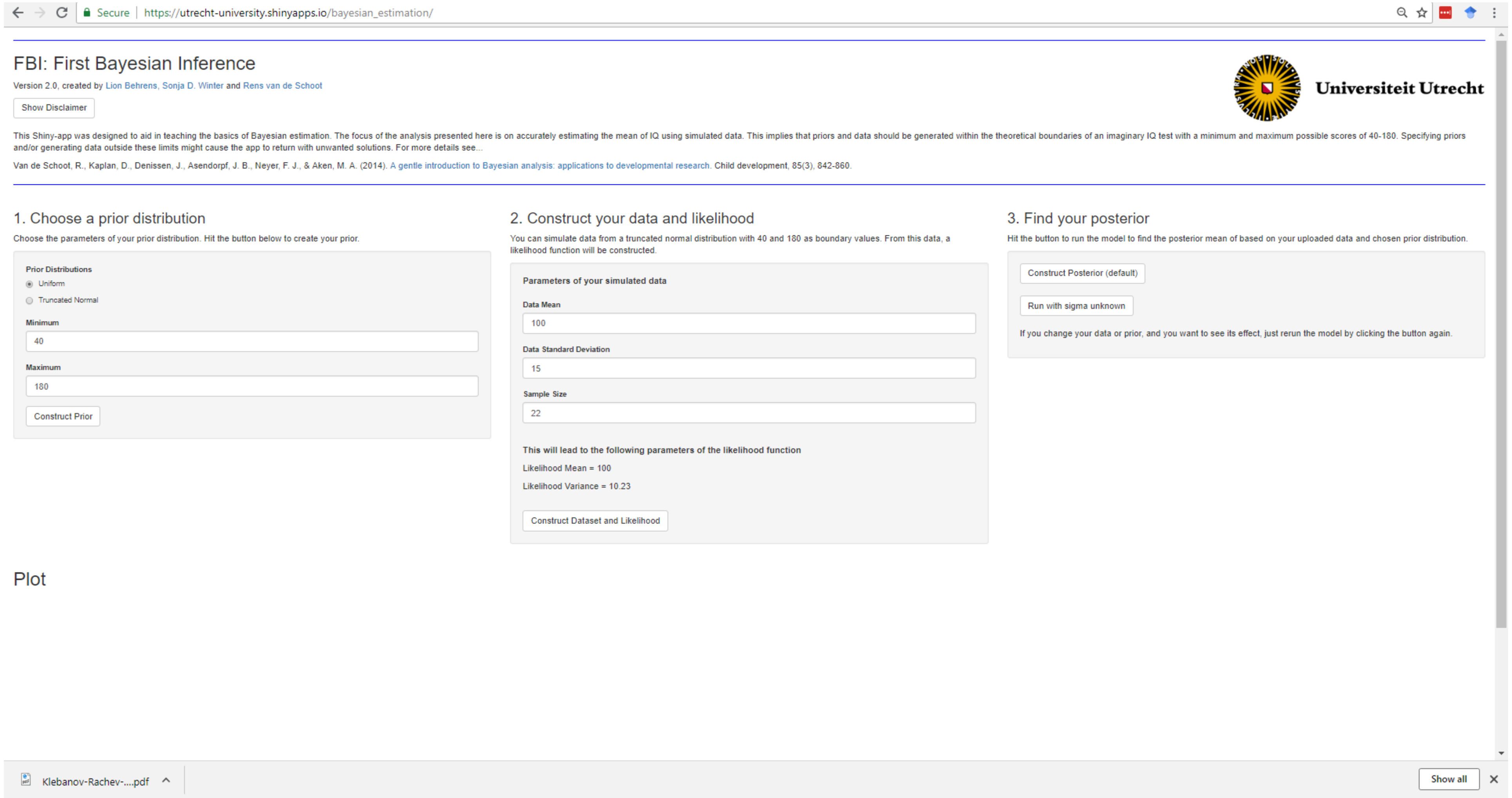
www.nature.com/articles/s43586-020-00001-2/


0.6
0.4
0.2
0.0
0
25
50
75
100
test
likelihood
posterior
prior
observation


0.4
0.2
0.0
0
25
50
75
100
test
likelihood
posterior
prior
observation

Exercise 1

Go to:
www.rensvandeschoot.com/FBI



1. Choose a prior distribution

Choose the parameters of your prior distribution. Hit the button below to create your prior.

Prior Distributions

- (•) Uniform
- ( ) Truncated Normal

Minimum

40

Maximum

180

Construct Prior

2. Construct your data and likelihood

You can simulate data from a truncated normal distribution with 40 and 180 as boundary values. From this data, a likelihood function will be constructed.

Parameters of your simulated data

Data Mean

100

Data Standard Deviation

15

Sample Size

20

This will lead to the following parameters of the likelihood function

Likelihood Mean = 100

Likelihood Variance = 11.25

Construct Dataset and Likelihood

3. Find your posterior

Hit the button to run the model to find the posterior mean of based on your uploaded data and chosen prior distribution.

Construct Posterior (default)

Run with sigma unknown

If you change your data or prior, and you want to see its effect, just rerun the model by clicking the button again.

1. Choose a prior distribution

Choose the parameters of your prior distribution. Hit the button below to create your prior.

Prior Distributions

- Uniform
- Truncated Normal

Minimum

40

Maximum

180

Construct Prior

2. Construct your data and likelihood

You can simulate data from a truncated normal distribution with 40 and 180 as boundary values. From this data, a likelihood function will be constructed.

Parameters of your simulated data

Data Mean

100

Data Standard Deviation

15

Sample Size

20

This will lead to the following parameters of the likelihood function

Likelihood Mean = 100

Likelihood Variance = 11.25

Construct Dataset and Likelihood

3. Find your posterior

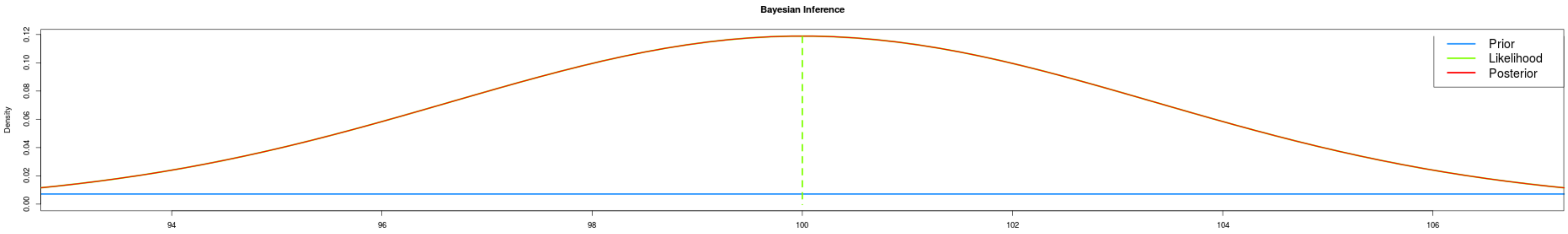
Hit the button to run the model to find the posterior mean of based on your uploaded data and chosen prior distribution.

Construct Posterior (default)

Run with sigma unknown

If you change your data or prior, and you want to see its effect, just rerun the model by clicking the button again.

Plot



1. Choose a prior distribution

Choose the parameters of your prior distribution. Hit the button below to create your prior.

Prior Distributions

- ( ) Uniform
- (•) Truncated Normal

Prior Mean

100

Prior Variance

10

Lower bound

40

Higher bound

180

Construct Prior

1. Choose a prior distribution

Choose the parameters of your prior distribution. Hit the button below to create your prior.

Prior Distributions

- ○ Uniform
- ◉ Truncated Normal

Prior Mean

100

Prior Variance

10

Lower bound

40

Higher bound

180

Construct Prior

2. Construct your data and likelihood

You can simulate data from a truncated normal distribution with 40 and 180 as boundary values. From this data, a likelihood function will be constructed.

Parameters of your simulated data

Data Mean

100

Data Standard Deviation

15

Sample Size

20

This will lead to the following parameters of the likelihood function

Likelihood Mean = 100

Likelihood Variance = 11.25

Construct Dataset and Likelihood

3. Find your posterior

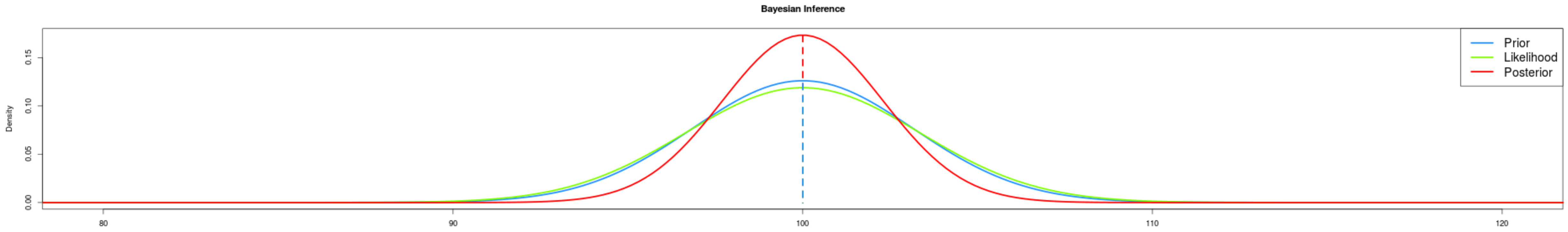
Hit the button to run the model to find the posterior mean of based on your uploaded data and chosen prior distribution.

Construct Posterior (default)

Run with sigma unknown

If you change your data or prior, and you want to see its effect, just rerun the model by clicking the button again.

Plot



1. Choose a prior distribution

Choose the parameters of your prior distribution. Hit the button below to create your prior.

Prior Distributions

- ○ Uniform
- ◉ Truncated Normal

Prior Mean

100

Prior Variance

2

Lower bound

40

Higher bound

180

Construct Prior

2. Construct your data and likelihood

You can simulate data from a truncated normal distribution with 40 and 180 as boundary values. From this data, a likelihood function will be constructed.

Parameters of your simulated data

Data Mean

100

Data Standard Deviation

15

Sample Size

20

This will lead to the following parameters of the likelihood function

Likelihood Mean = 100

Likelihood Variance = 11.25

Construct Dataset and Likelihood

3. Find your posterior

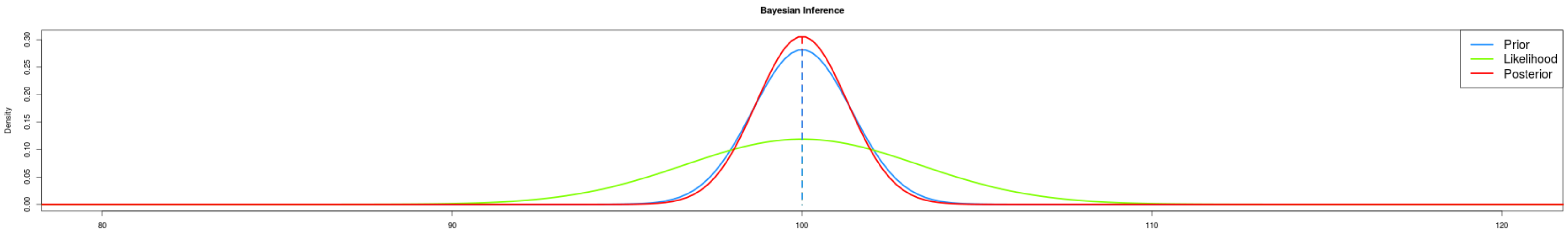
Hit the button to run the model to find the posterior mean of based on your uploaded data and chosen prior distribution.

Construct Posterior (default)

Run with sigma unknown

If you change your data or prior, and you want to see its effect, just rerun the model by clicking the button again.

Plot



1. Choose a prior distribution

Choose the parameters of your prior distribution. Hit the button below to create your prior.

Prior Distributions

- Uniform
- Truncated Normal

Prior Mean

90

Prior Variance

10

Lower bound

40

Higher bound

180

Construct Prior

2. Construct your data and likelihood

You can simulate data from a truncated normal distribution with 40 and 180 as boundary values. From this data, a likelihood function will be constructed.

Parameters of your simulated data

Data Mean

100

Data Standard Deviation

15

Sample Size

20

This will lead to the following parameters of the likelihood function

Likelihood Mean = 100

Likelihood Variance = 11.25

Construct Dataset and Likelihood

3. Find your posterior

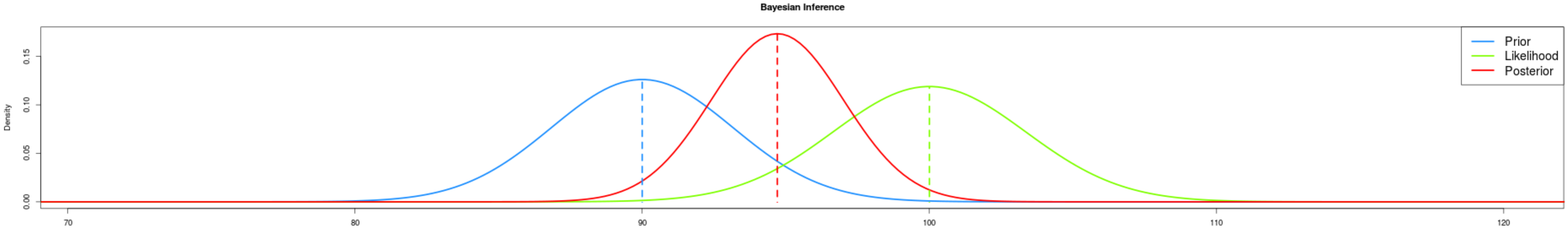
Hit the button to run the model to find the posterior mean of based on your uploaded data and chosen prior distribution.

Construct Posterior (default)

Run with sigma unknown

If you change your data or prior, and you want to see its effect, just rerun the model by clicking the button again.

Plot



1. Choose a prior distribution

Choose the parameters of your prior distribution. Hit the button below to create your prior.

Prior Distributions

- Uniform
- Truncated Normal

Prior Mean

70

Prior Variance

10

Lower bound

40

Higher bound

180

Construct Prior

2. Construct your data and likelihood

You can simulate data from a truncated normal distribution with 40 and 180 as boundary values. From this data, a likelihood function will be constructed.

Parameters of your simulated data

Data Mean

100

Data Standard Deviation

15

Sample Size

20

This will lead to the following parameters of the likelihood function

Likelihood Mean = 100

Likelihood Variance = 11.25

Construct Dataset and Likelihood

3. Find your posterior

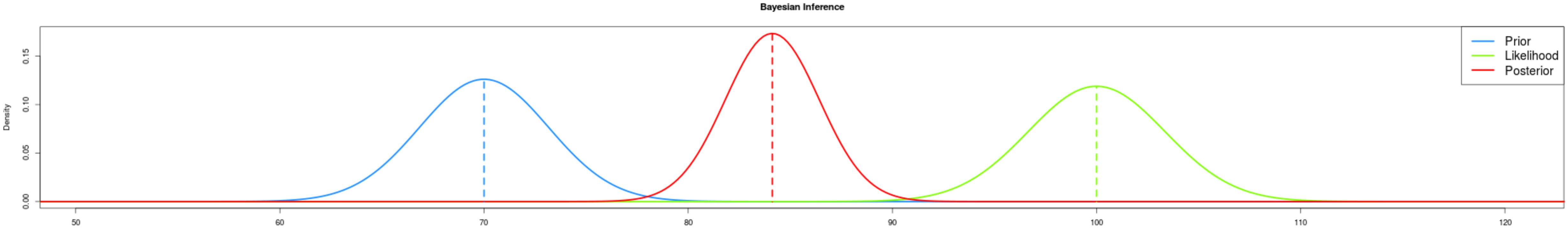
Hit the button to run the model to find the posterior mean of based on your uploaded data and chosen prior distribution.

Construct Posterior (default)

Run with sigma unknown

If you change your data or prior, and you want to see its effect, just rerun the model by clicking the button again.

Plot



3. Find your posterior

Hit the button to run the model to find the posterior mean of based on your uploaded data and chosen prior distribution.

Construct Posterior (default)

Run with sigma unknown

If you change your data or prior, and you want to see its effect, just rerun the model by clicking the button again.

Conjugate priors with fixed parameters

When likelihood function is a continuous distribution [edit]

Likelihood	Model parameters	Conjugate prior distribution	Prior hyperparameters	Posterior hyperparameters	Interpretation of hyperparameters	Posterior predictive[note 4]
Normal with known variance σ^2	μ (mean)	Normal	$\mu_0,\ \sigma_0^2$	$\frac{1}{\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}}\left(\frac{\mu_0}{\sigma_0^2}+\frac{\sum_{i=1}^n x_i}{\sigma^2}\right),\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)^{-1}$	mean was estimated from observations with total precision (sum of all individual precisions) $1/\sigma_0^2$ and with sample mean μ_0	$\mathcal{N}(\tilde{x}\vert\mu_0', \sigma_0^{2\prime}+\sigma^2)$[5]
Normal with known precision τ	μ (mean)	Normal	$\mu_0,\ \tau_0$	$\frac{\tau_0\mu_0+\tau\sum_{i=1}^n x_i}{\tau_0+n\tau},\ \tau_0+n\tau$	mean was estimated from observations with total precision (sum of all individual precisions) τ_0 and with sample mean μ_0	$\mathcal{N}\left(\tilde{x}\mid\mu_0', \frac{1}{\tau_0'}+\frac{1}{\tau}\right)$[5]
Normal with known mean μ	σ^2 (variance)	Inverse gamma	$\alpha,\ \beta$ [note 5]	$\alpha+\frac{n}{2},\ \beta+\frac{\sum_{i=1}^n(x_i-\mu)^2}{2}$	variance was estimated from 2α observations with sample variance β/α (i.e. with sum of squared deviations 2β, where deviations are from known mean μ)	$t_{2\alpha'}(\tilde{x}\vert\mu, \sigma^2=\beta'/\alpha')$[5]
Normal with known mean μ	σ^2 (variance)	Scaled inverse chi-squared	$\nu,\ \sigma_0^2$	$\nu+n,\ \frac{\nu\sigma_0^2+\sum_{i=1}^n(x_i-\mu)^2}{\nu+n}$	variance was estimated from ν observations with sample variance σ_0^2	$t_{\nu'}(\tilde{x}\vert\mu, \sigma_0^{2\prime})$[5]
Normal with known mean μ	τ (precision)	Gamma	$\alpha,\ \beta$[note 3]	$\alpha+\frac{n}{2},\ \beta+\frac{\sum_{i=1}^n(x_i-\mu)^2}{2}$	precision was estimated from 2α observations with sample variance β/α (i.e. with sum of squared deviations 2β, where deviations are from known mean μ)	$t_{2\alpha'}(\tilde{x}\mid\mu, \sigma^2=\beta'/\alpha')$[5]
Normal[note 6]	μ and σ^2 Assuming exchangeability	Normal-inverse gamma	$\mu_0,\ \nu,\ \alpha,\ \beta$	$\frac{\nu\mu_0+n\bar{x}}{\nu+n},\ \nu+n,\ \alpha+\frac{n}{2},$ $\beta+\frac{1}{2}\sum_{i=1}^n(x_i-\bar{x})^2+\frac{n\nu}{\nu+n}\frac{(\bar{x}-\mu_0)^2}{2}$ • $\bar{x}$ is the sample mean	mean was estimated from ν observations with sample mean μ_0; variance was estimated from 2α observations with sample mean μ_0 and sum of squared deviations 2β	$t_{2\alpha'}\left(\tilde{x}\mid\mu', \frac{\beta'(\nu'+1)}{\nu'\alpha'}\right)$[5]
Normal	μ and τ Assuming exchangeability	Normal-gamma	$\mu_0,\ \nu,\ \alpha,\ \beta$	$\frac{\nu\mu_0+n\bar{x}}{\nu+n},\ \nu+n,\ \alpha+\frac{n}{2},$ $\beta+\frac{1}{2}\sum_{i=1}^n(x_i-\bar{x})^2+\frac{n\nu}{\nu+n}\frac{(\bar{x}-\mu_0)^2}{2}$ • $\bar{x}$ is the sample mean	mean was estimated from ν observations with sample mean μ_0, and precision was estimated from 2α observations with sample mean μ_0 and sum of squared deviations 2β	$t_{2\alpha'}\left(\tilde{x}\mid\mu', \frac{\beta'(\nu'+1)}{\alpha'\nu'}\right)$[5]

https://en.wikipedia.org/wiki/Conjugate_prior

How to obtain posterior?

In complex models, the posterior is often intractable (impossible to compute exactly)

Solution: approximate posterior by simulation

Simulate many draws from posterior distribution
Compute mode, median, mean, 95% interval et cetera from the simulated draws

ANOVA example

4 unknown parameters μ_j (j=1,...,4) and one common but unknown σ^2.

Statistical model:

$$Y = I + \mu_1 * D_1 + \mu_2 * D_2 + \mu_3 * D_3 + E$$

with $E \sim N(0, \sigma^2)$

ANOVA example

4 unknown parameters μ_j ($j=1,\ldots,4$) and one common but unknown σ^2.

Statistical model:

$$Y = \mu_1 * D_1 + \mu_2 * D_2 + \mu_3 * D_3 + \mu_4 * D_4 + E$$

with $E \sim N(0, \sigma^2)$

Priors

Specify prior: $\Pr(\mu_1, \mu_2, \mu_3, \mu_4, \sigma^2)$

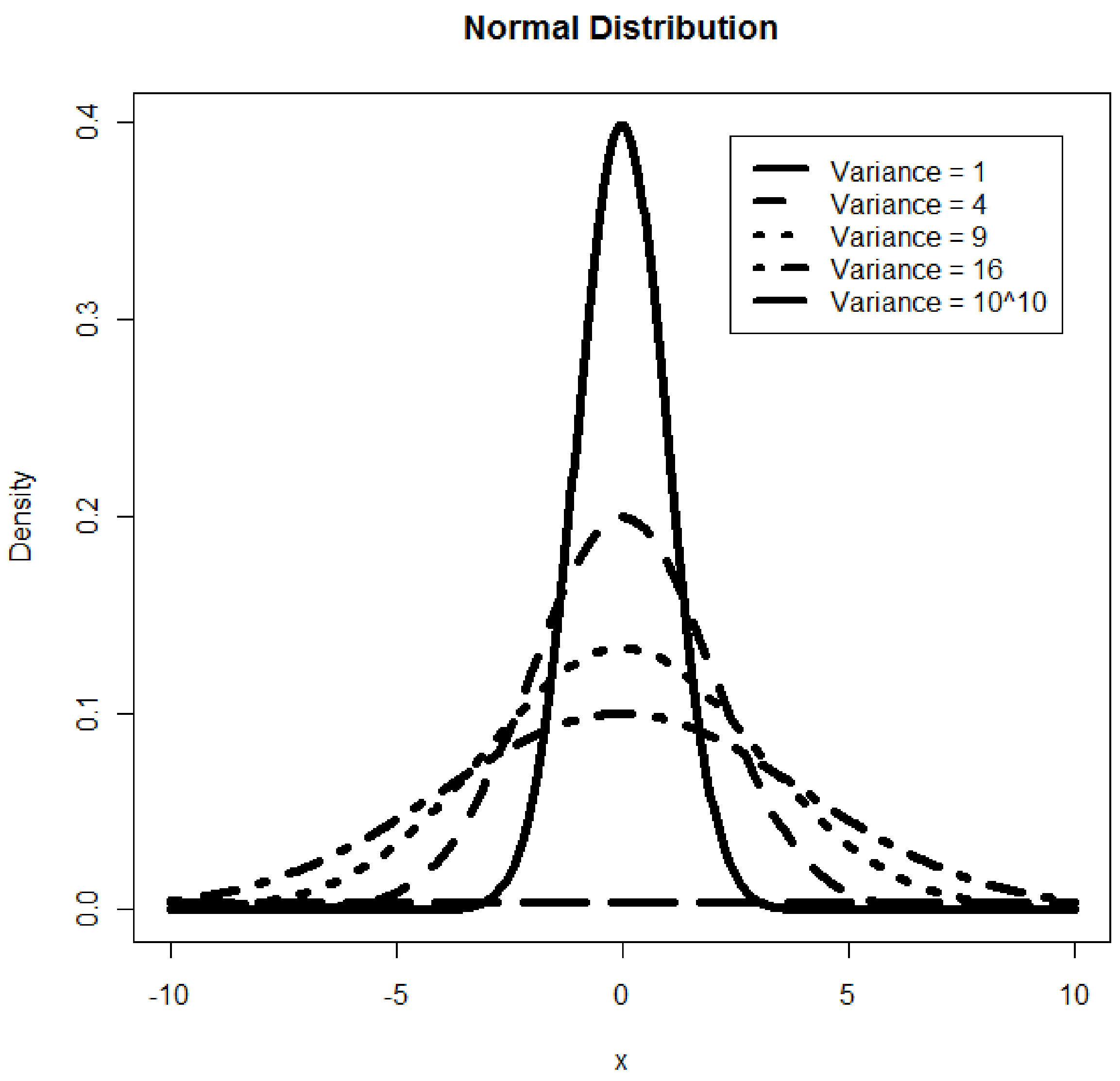
Priors

Specify prior: $\Pr(\mu_1, \mu_2, \mu_3, \mu_4, \sigma^2)$

Prior $(\mu_j) \sim \mathrm{Nor}(\mu_0, var_0)$

Prior $(\mu_j) \sim \mathrm{Nor}(0{,}10000)$

Hyperparameters:
μ (mean), σ^2 (variance)



Priors

Specify prior: $Pr(\mu_1, \mu_2, \mu_3, \mu_4, \sigma^2)$

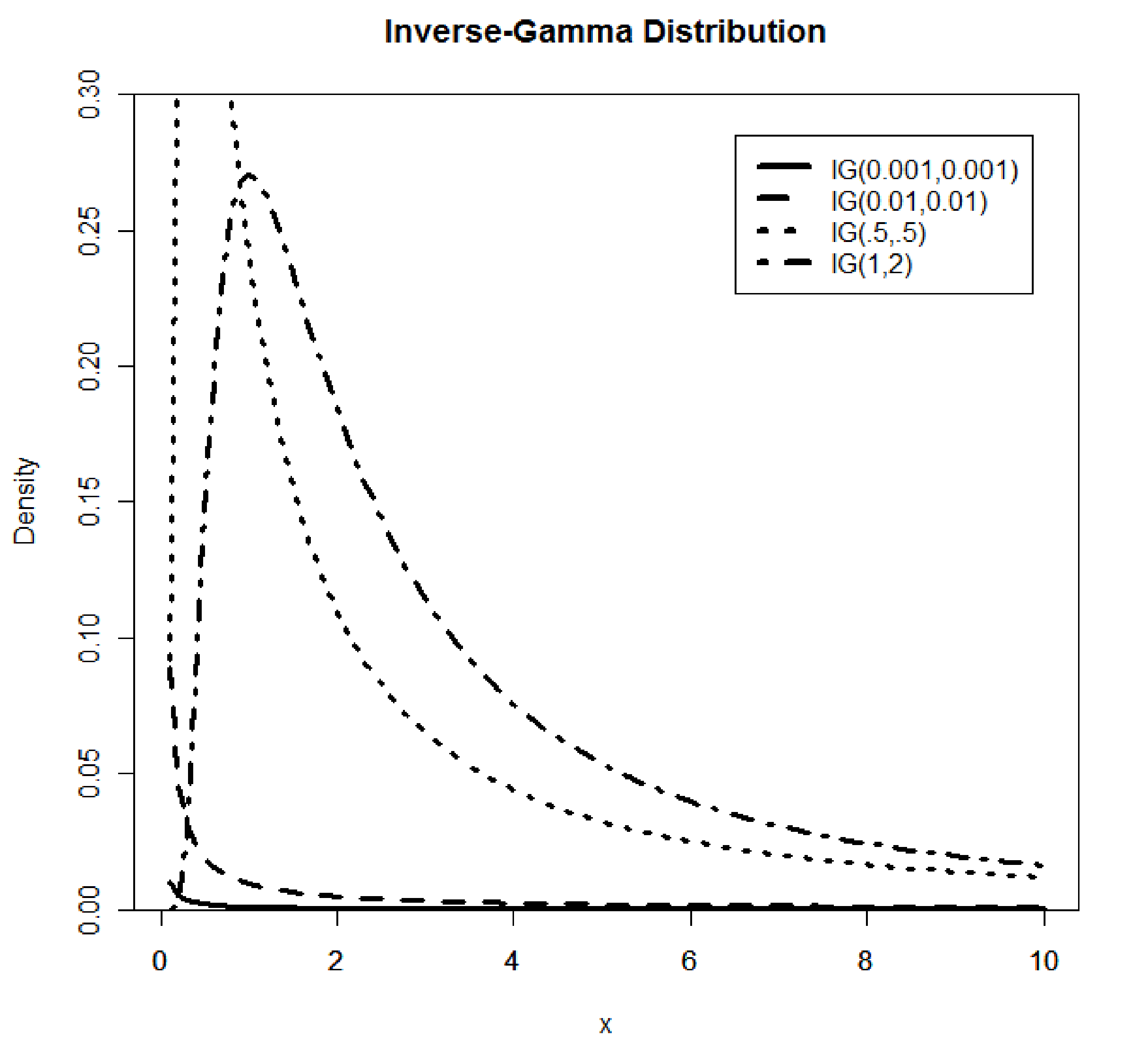
Prior (μ_j) ~ Nor(μ_0, var_0)

Prior (μ_j) ~ Nor(0,10000)

Prior (σ^2) ~ IG(0.001, 0.001)

Universiteit Utrecht

Hyperparameters:
α (shape), β (scale)



Posterior

Combine prior with likelihood provides posterior:

Post (μ_1, μ_2, μ_3, μ_4, σ^2 | data)

…this is a 5 dimensional distribution…

The Gibbs sampler

Iterative evaluation via conditional distributions:

Post (μ_1 | μ_2, μ_3, μ_4, σ^2, data) ~ *Prior* (μ_1) *X Data* (μ_1)

Post (μ_2 | μ_1, μ_3, μ_4, σ^2, data) ~ *Prior* (μ_2) *X Data* (μ_2)

Post (μ_3 | μ_1, μ_2, μ_4, σ^2, data) ~ *Prior* (μ_3) *X Data* (μ_3)

Post (μ_4 | μ_1, μ_2, μ_3, σ^2, data) ~ *Prior* (μ_4) *X Data* (μ_4)

Post (σ^2 | μ_1, μ_2, μ_3, μ_4, data) ~ *Prior* (σ^2) *X Data* (σ^2)

The Gibbs sampler

1. Assign starting values
2. Sample μ_1 from conditional distribution
3. Sample μ_2 from conditional distribution
4. Sample μ_3 from conditional distribution
5. Sample μ_4 from conditional distribution
6. Sample σ^2 from conditional distribution
7. Go to step 2 over and over again

$$\mu_1 * D_1 + \mu_2 * D_2 + \mu_3 * D_3 + \mu_4 * D_4 + E$$

Step 1: assign starting values

Step 1: $3*D_1 + 5*D_2 + 8*D_3 + 3*D_4 + 10$

Step 2: Sample μ_1 from conditional distribution

Step 1: $3*D_1 + 5*D_2 + 8*D_3 + 3*D_4 + 10$

Step 2:  $+ \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 2: Sample μ_1 from conditional distribution

Step 1: $3*D_1 + 5*D_2 + 8*D_3 + 3*D_4 + 10$

Step 2: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 2: Sample μ_1 from conditional distribution

Step 1: $3*D_1 + 5*D_2 + 8*D_3 + 3*D_4 + 10$

Step 2: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 3: Sample μ_2 from conditional distribution

Step 1: $3*D_1 + 5*D_2 + 8*D_3 + 3*D_4 + 10$

Step 2: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 3: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 3: Sample μ_2 from conditional distribution

Step 1: $3*D_1 + 5*D_2 + 8*D_3 + 3*D_4 + 10$

Step 2: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 3: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Do this for all parameters

Step 1: $3*D_1 + 5*D_2 + 8*D_3 + 3*D_4 + 10$

Step 2: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 3: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 4: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 5: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 6: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

This is iteration 1

Step 1: $3*D_1 + 5*D_2 + 8*D_3 + 3*D_4 + 10$

Iteration 1

Step 2: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 3: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 4: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 5: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 6: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Replace starting values with new estimates

Step 1: $3*D_1 + 5*D_2 + 8*D_3 + 3*D_4 + 10$

Step 2: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 3: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 4: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 5: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 6: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 7: Go to step 2 and start with iteration 2

Step 1: $3*D_1 + 5*D_2 + 8*D_3 + 3*D_4 + 10$

Iteration 2

Step 2: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 3: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 4: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 5: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 6: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Universiteit Utrecht

Repeat k times

Step 1: $3*D_1 + 5*D_2 + 8*D_3 + 3*D_4 + 10$

Iteration k

Step 2: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 3: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 4: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 5: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 6: $\mu_1*D_1 + \mu_2*D_2 + \mu_3*D_3 + \mu_4*D_4 + E$

Step 1: assign starting values

Iteration	μ_1	μ_2	μ_3	μ_4	σ^2
1	3.00	5.00	8.00	3.00	10

Step 2: Sample μ_1 from conditional distribution

Iteration	μ_1	μ_2	μ_3	μ_4	σ^2
1	3.00	5.00	8.00	3.00	10
2	3.75				

Step 3: Sample μ_2 from conditional distribution

Iteration	μ_1	μ_2	μ_3	μ_4	σ^2
1	3.00	5.00	8.00	3.00	10
2	3.75	4.25			

Step 6: Sample σ^2 from conditional distribution

Iteration	μ_1	μ_2	μ_3	μ_4	σ^2
1	3.00	5.00	8.00	3.00	10
2	3.75	4.25	7.00	4.30	8

Universiteit Utrecht

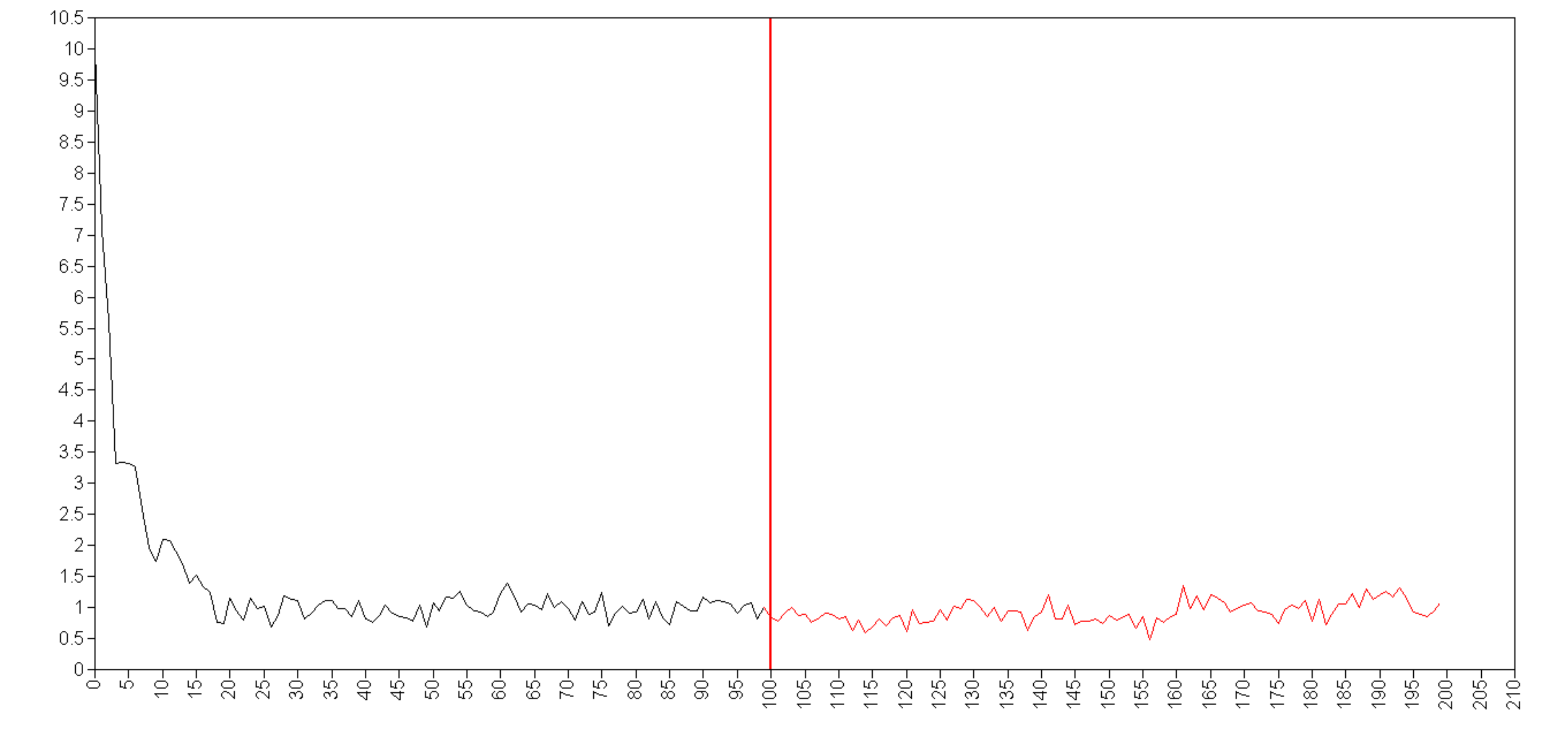
Step 7: Go to step 2 over and over again

Iteration	μ_1	μ_2	μ_3	μ_4	σ^2
1	3.00	5.00	8.00	3.00	10
2	3.75	4.25	7.00	4.30	8
3	3.65				

Step 7: Go to step 2 over and over again

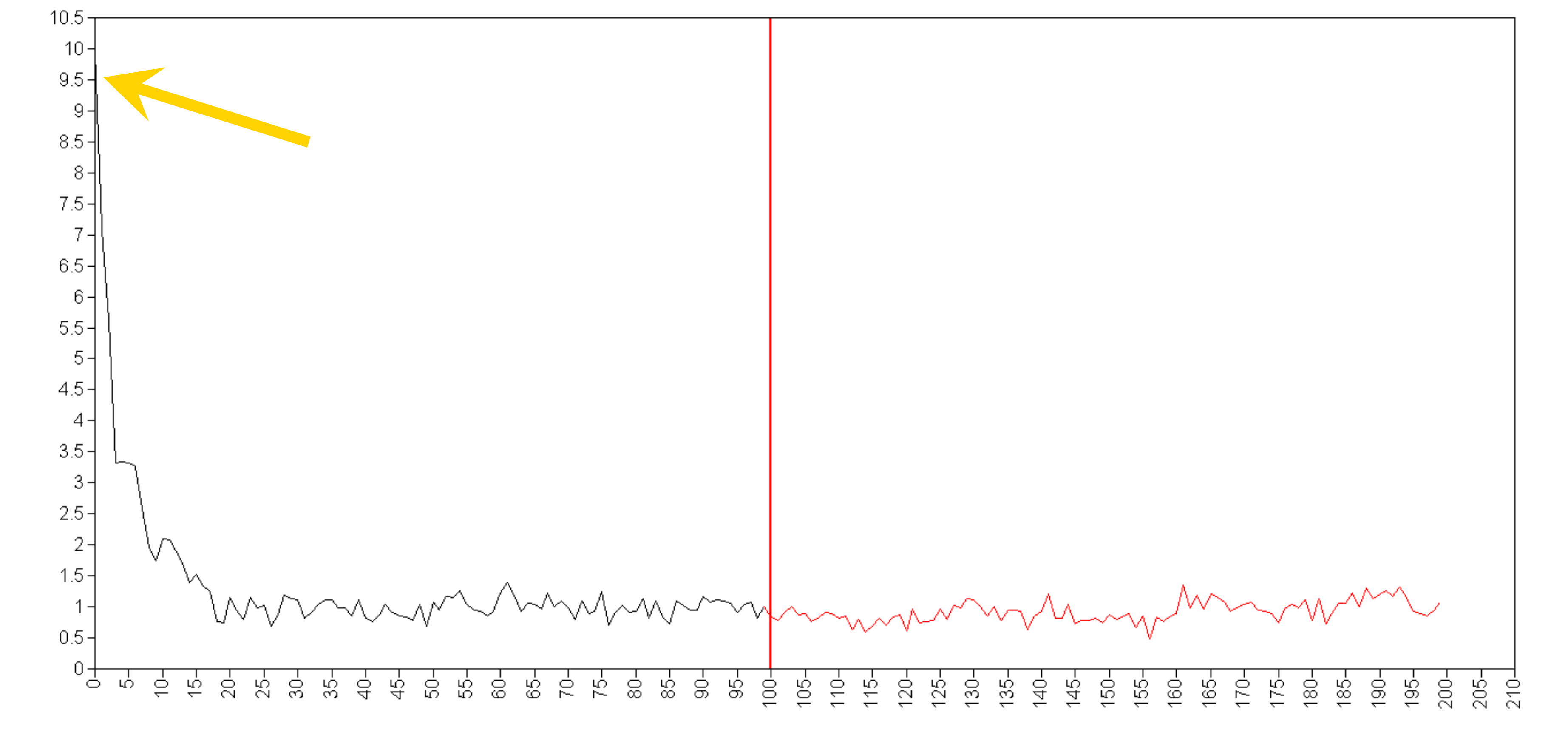
Iteration	μ_1	μ_2	μ_3	μ_4	σ^2
1	3.00	5.00	8.00	3.00	10
2	3.75	4.25	7.00	4.30	8
3	3.65	4.11	6.78	5.55	5
.	.	.	.	.	.
15	4.45	3.19	5.08	6.55	1.1
.	.	.	.	.	.
.	.	.	.	.	.
199	4.59	3.75	5.21	6.36	1.2
200	4.36	3.45	4.65	6.99	1.3

Trace plot

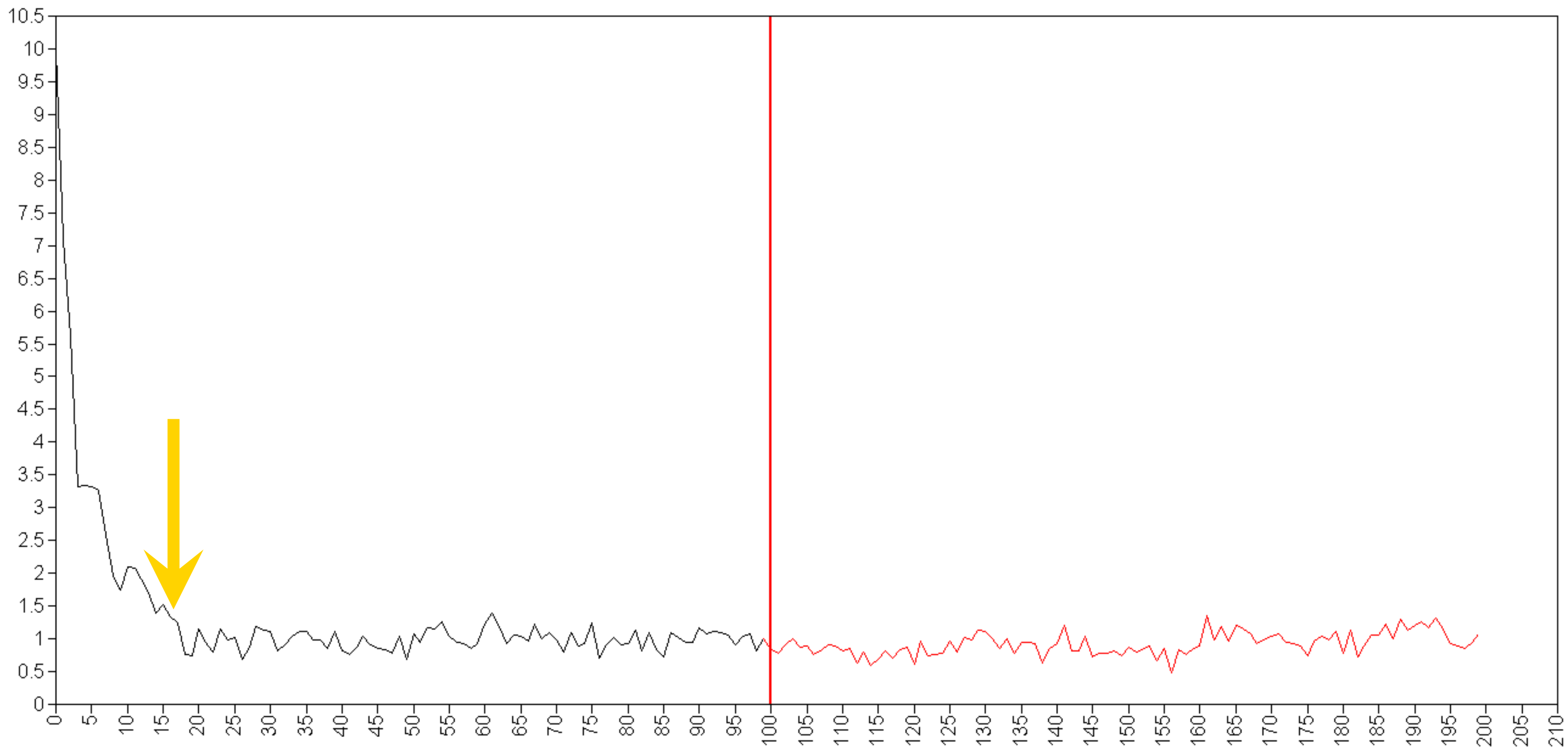


10.5
10
9.5
9
8.5
8
7.5
7
6.5
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210

Trace plot, starting value

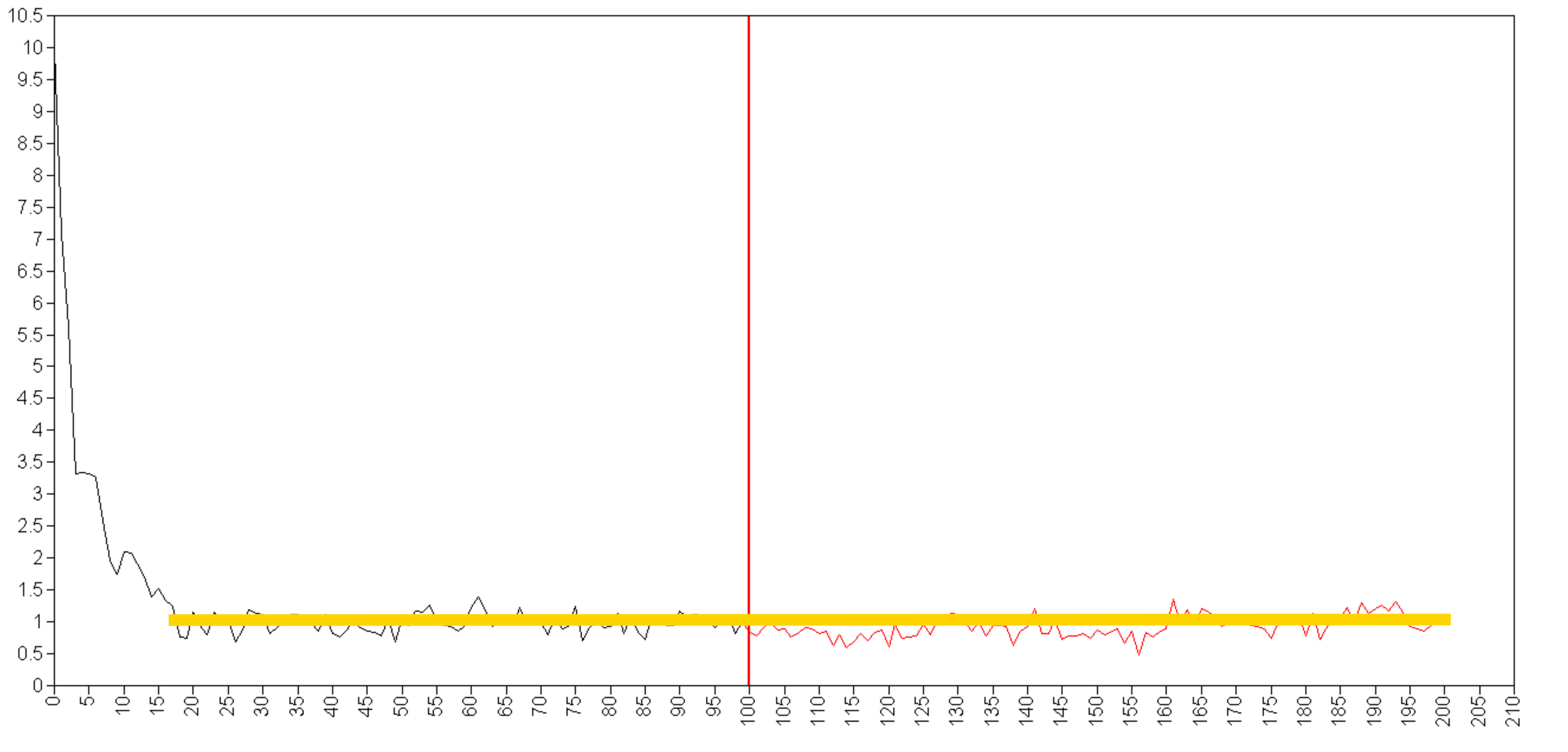


Trace plot

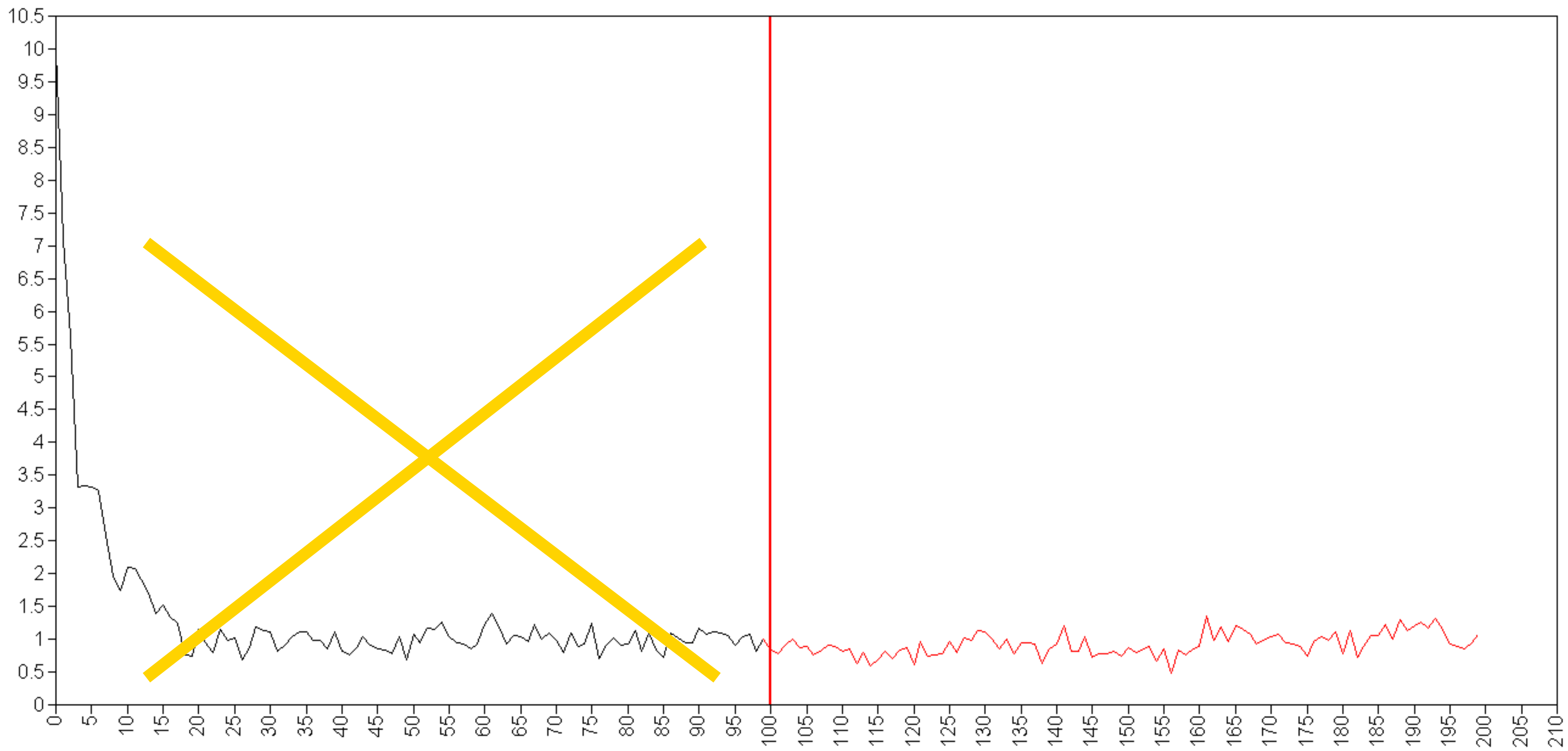

10.5
10
9.5
9
8.5
8
7.5
7
6.5
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210

Universiteit Utrecht

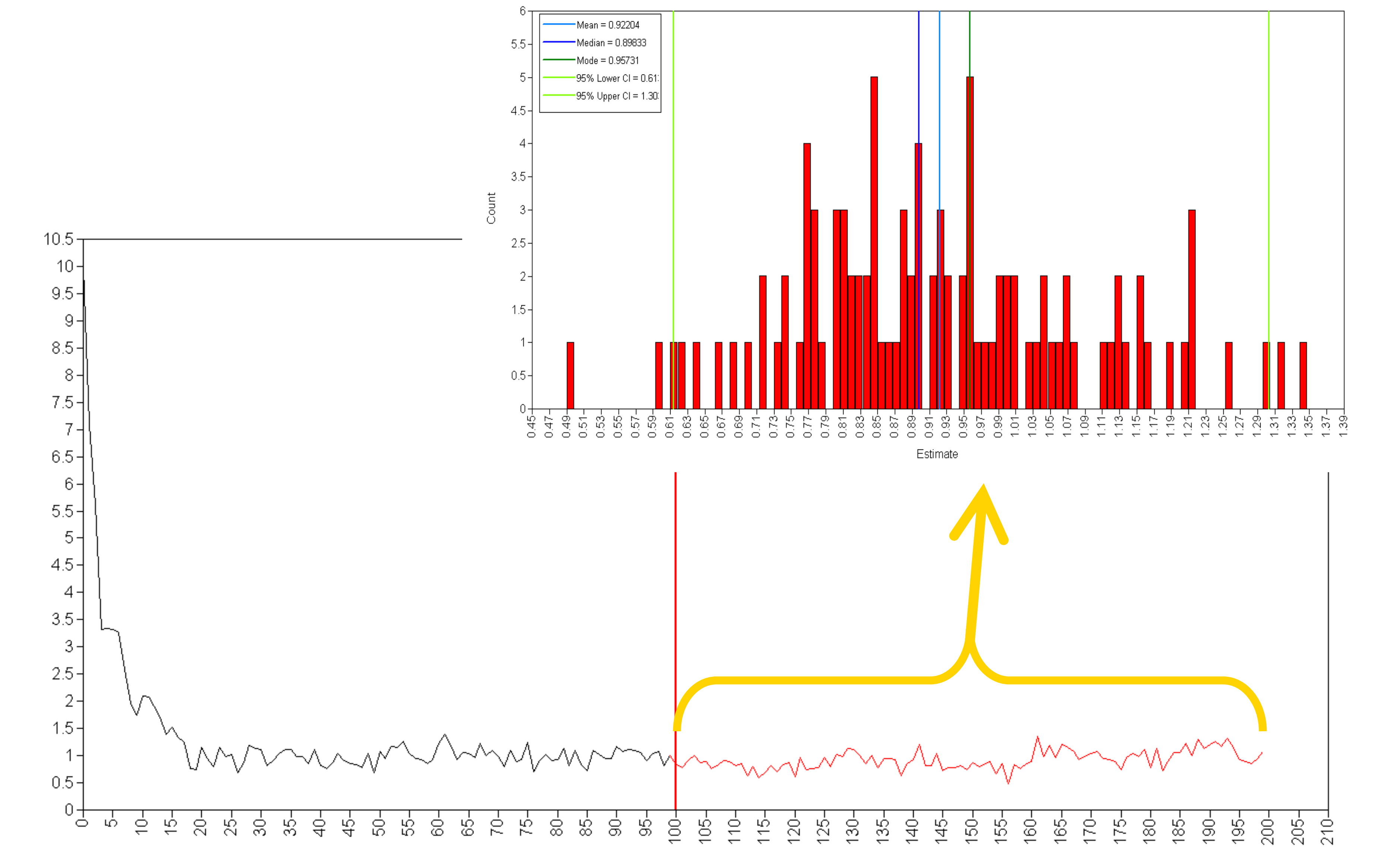
Trace plot, stationary distribution



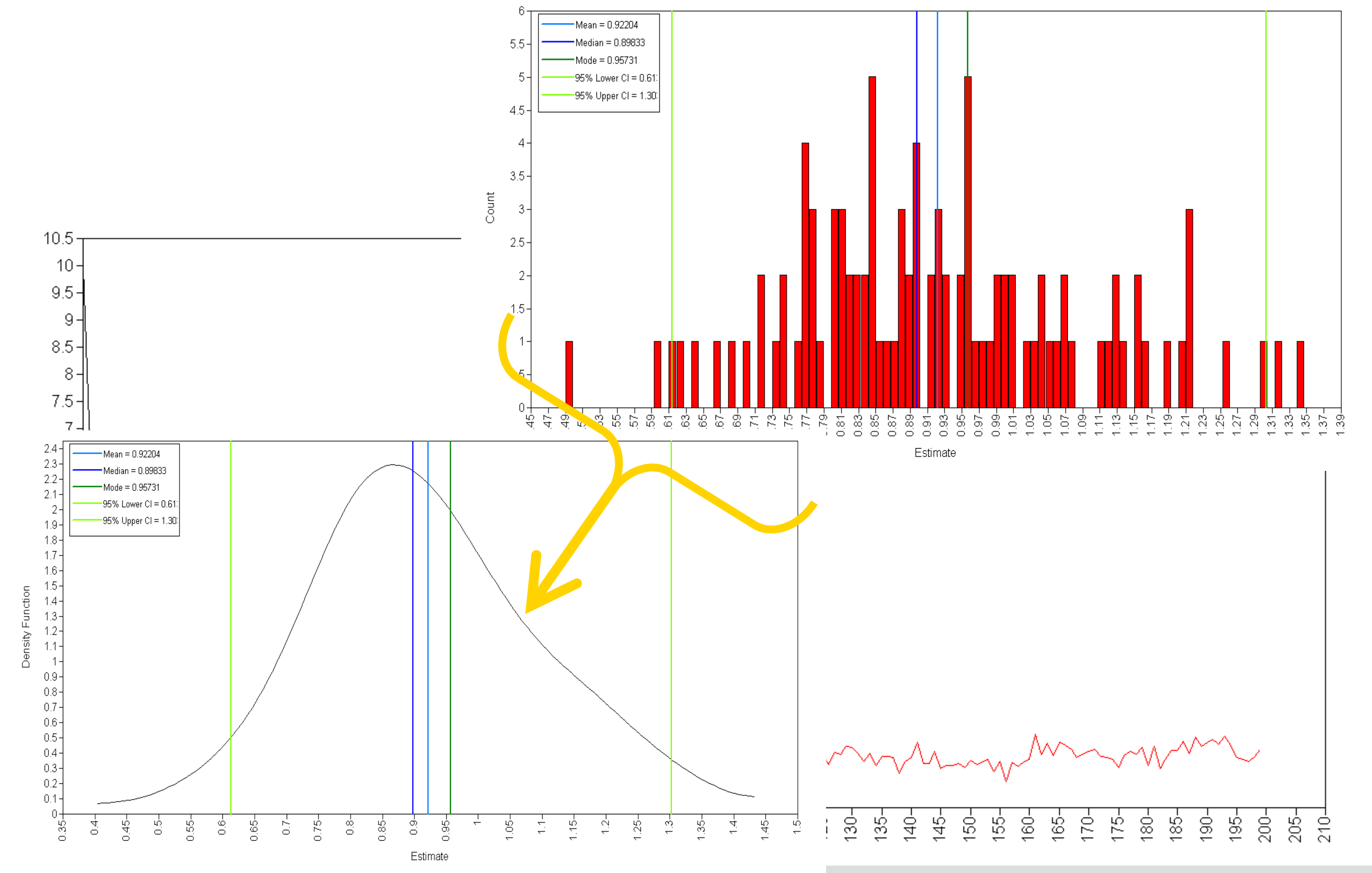
Trace plot, burn-in



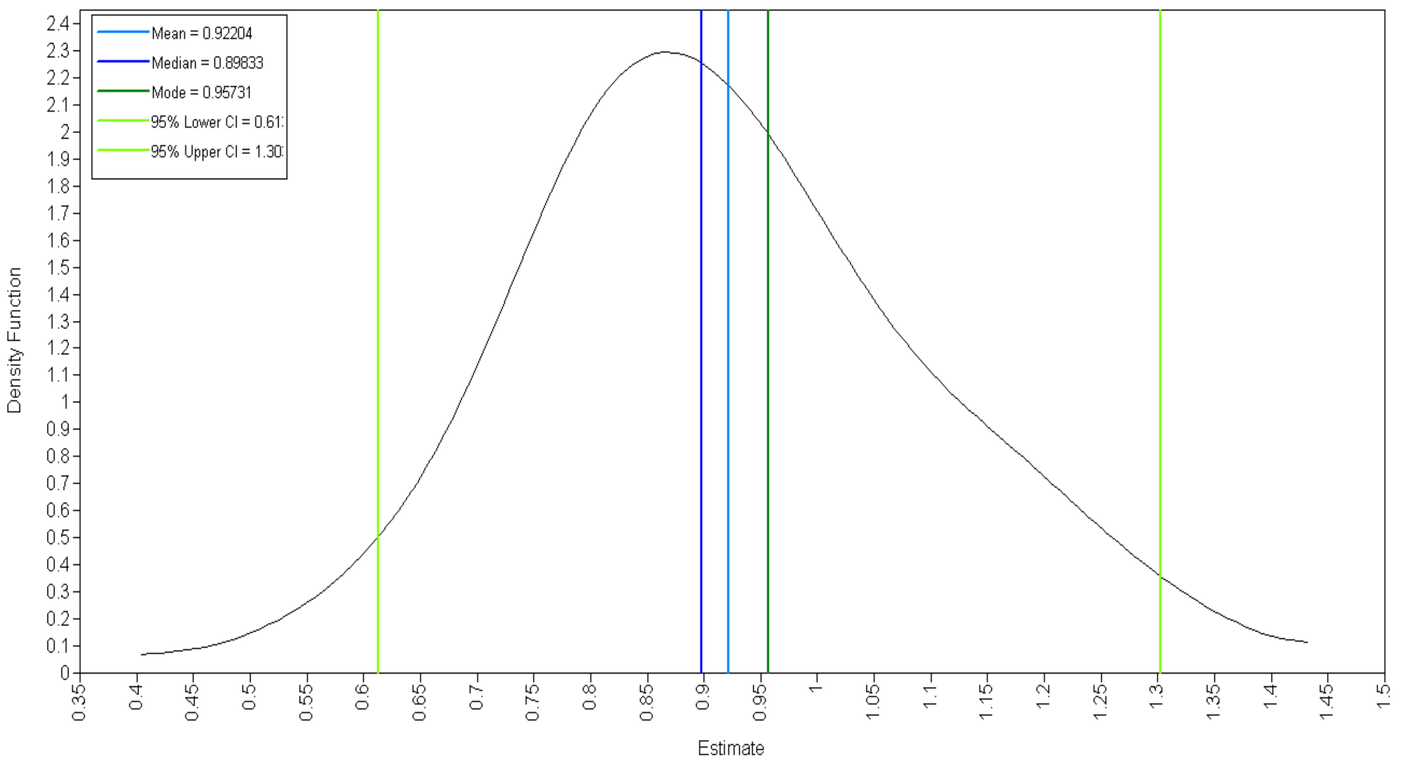
Histogram


Mean = 0.92204
Median = 0.89833
Mode = 0.95731
95% Lower CI = 0.61
95% Upper CI = 1.30
Count
Estimate

Kernel density plot

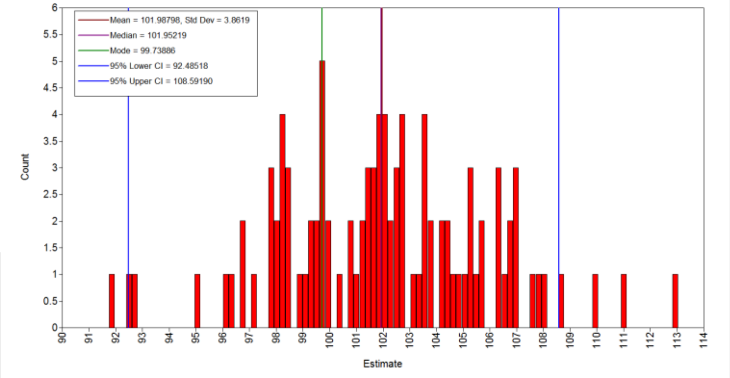


Approximation of the posterior distribution

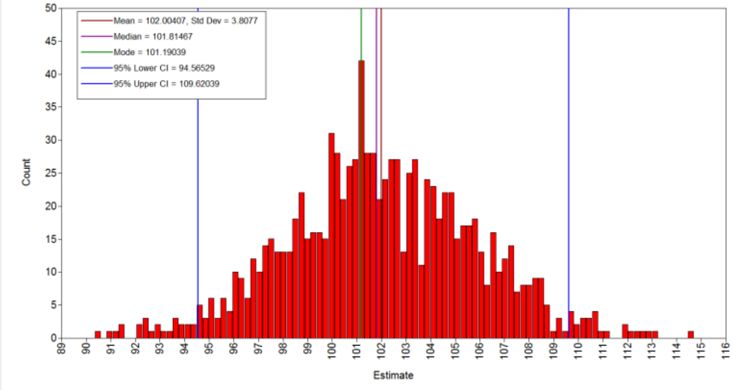


Universiteit Utrecht

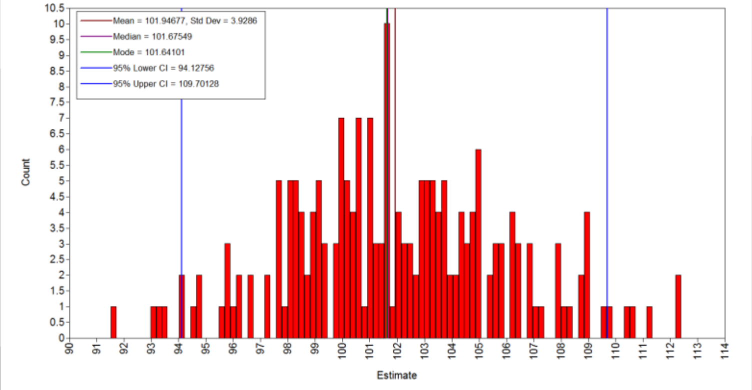
The more iterations, the more information in the histogram and the better the results approximate the posterior distribution



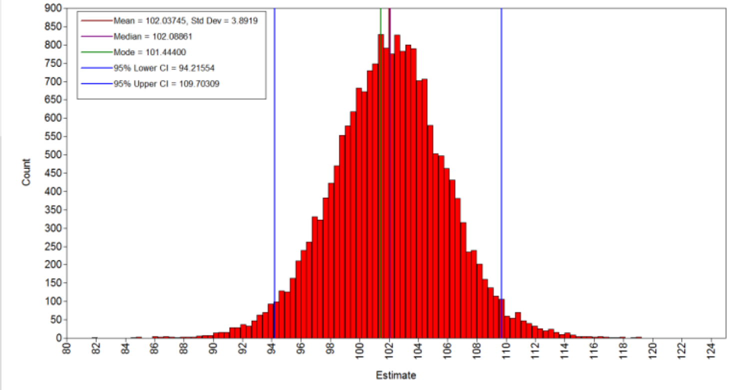
(A)



(C)

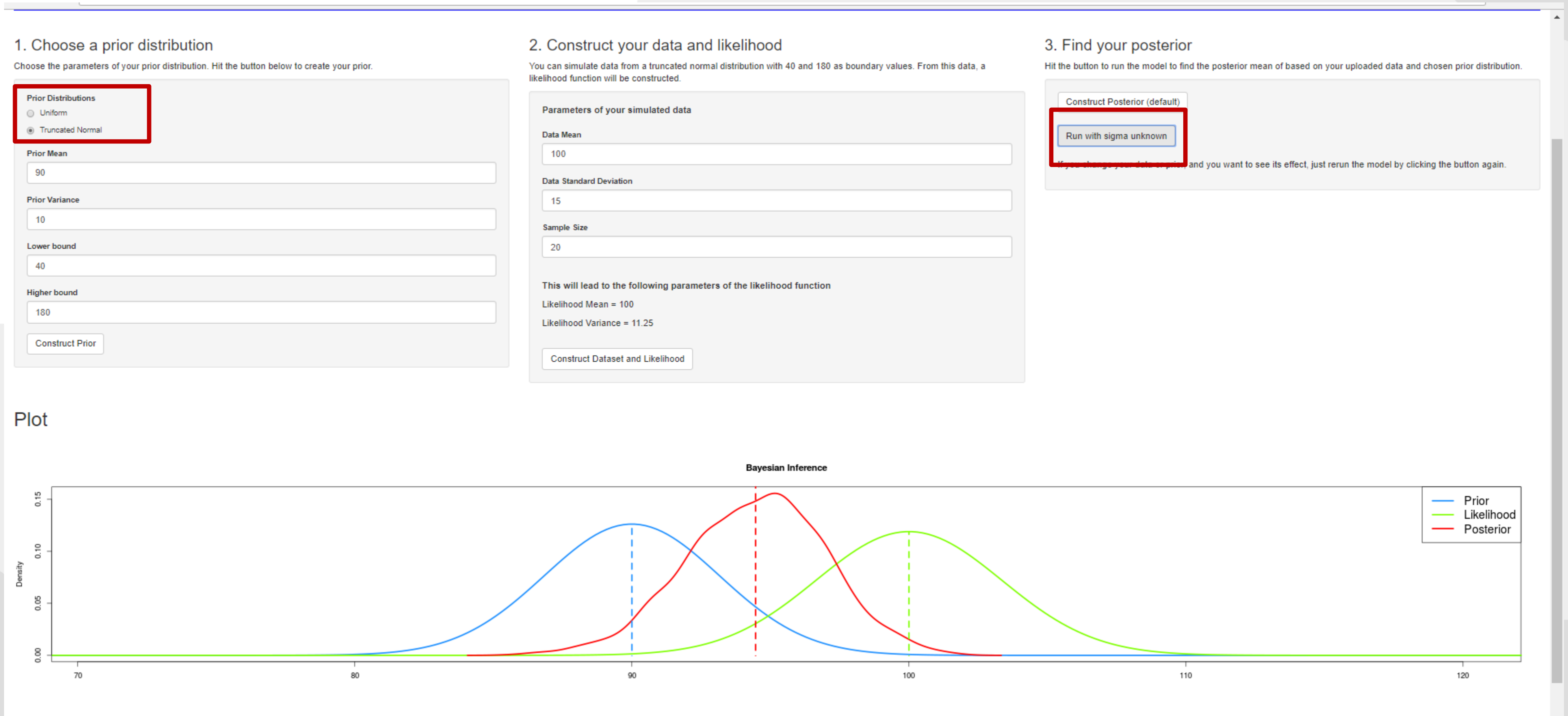


(B)



(D)

Universiteit Utrecht



Convergence

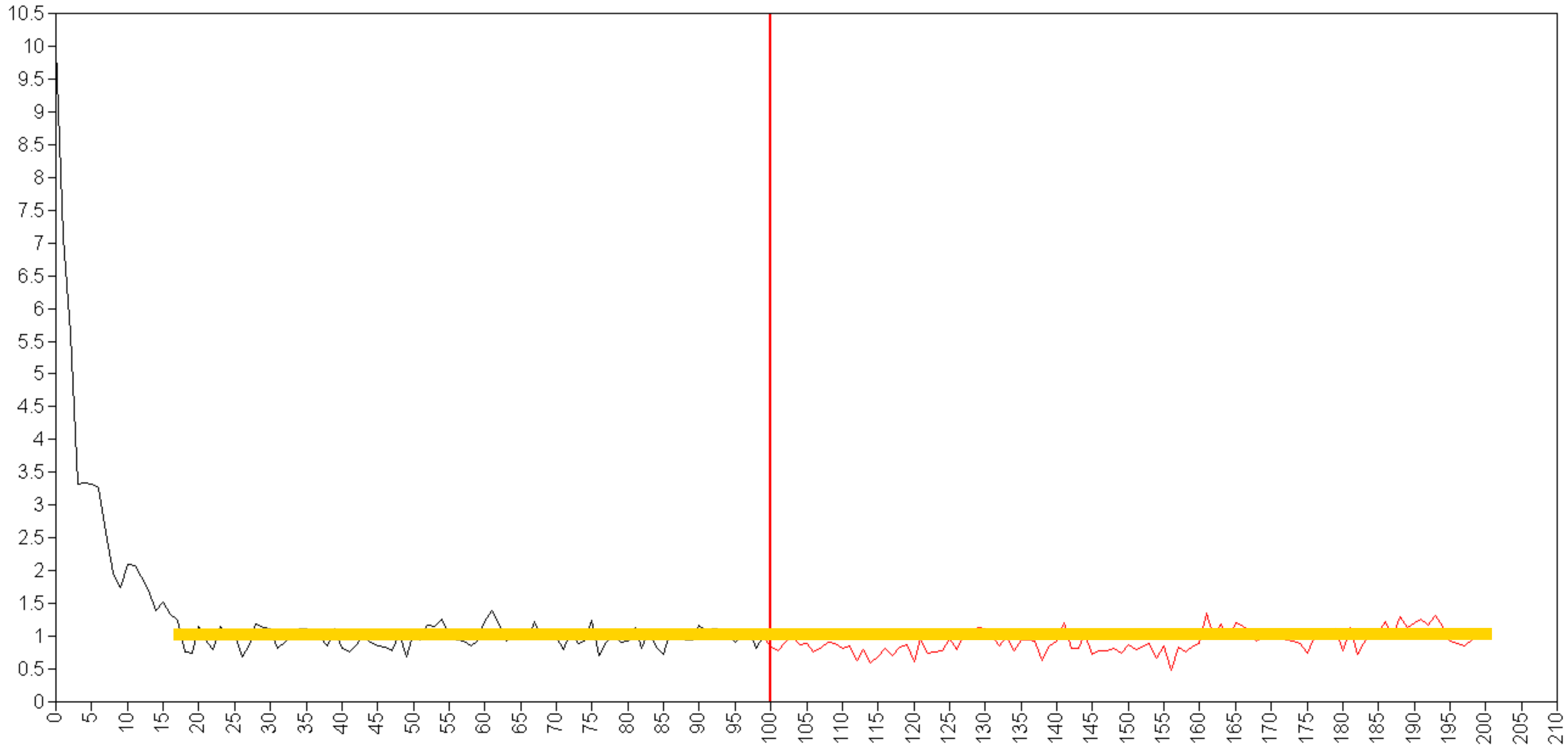
Sampler must run *t* iterations 'burn in' before we reach target distribution *f(Z)*

- How many iterations are needed to converge on the target distribution?

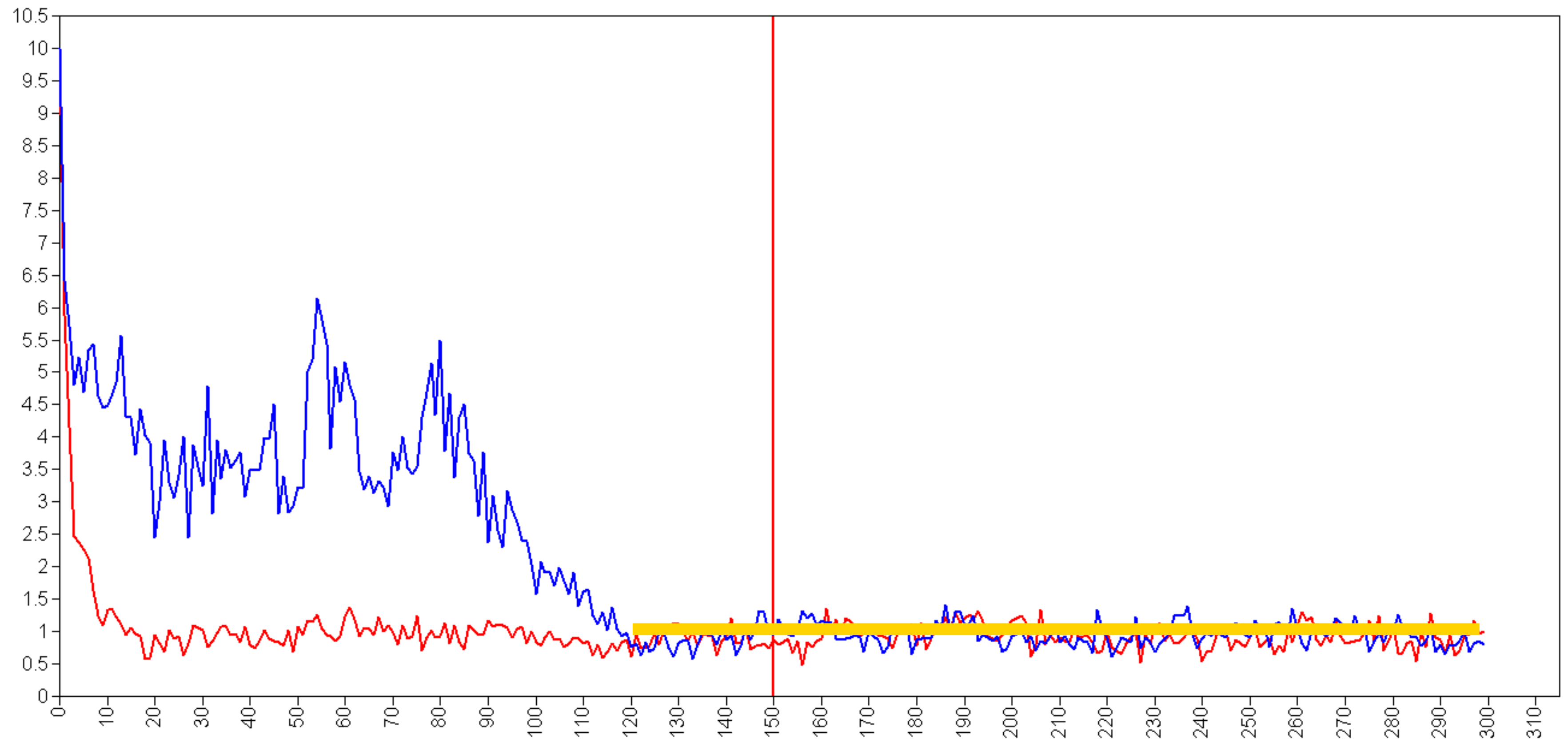
Diagnostics

- Examine graph of burn in
- Try different starting values
- Run several chains in parallel

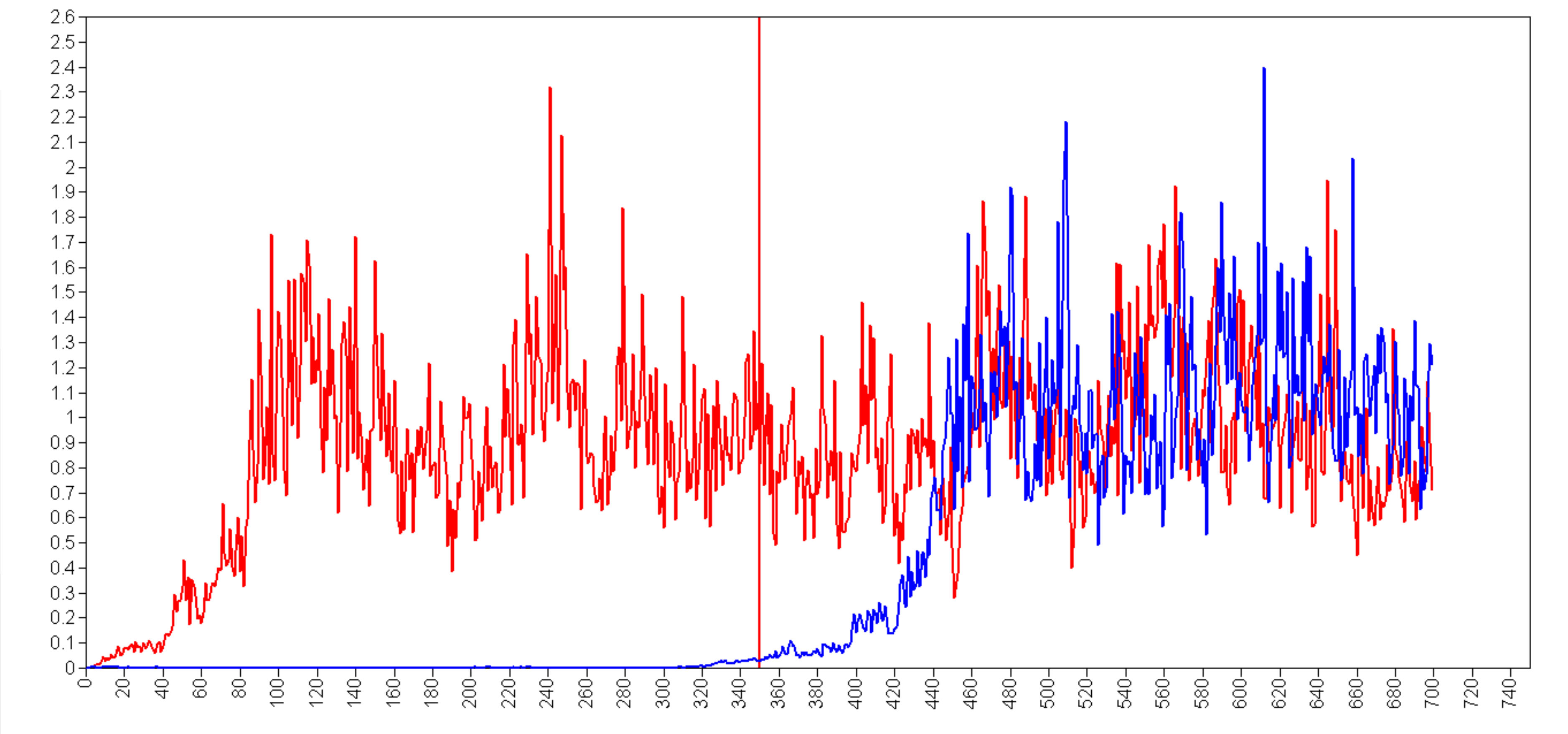
Convergence



Convergence

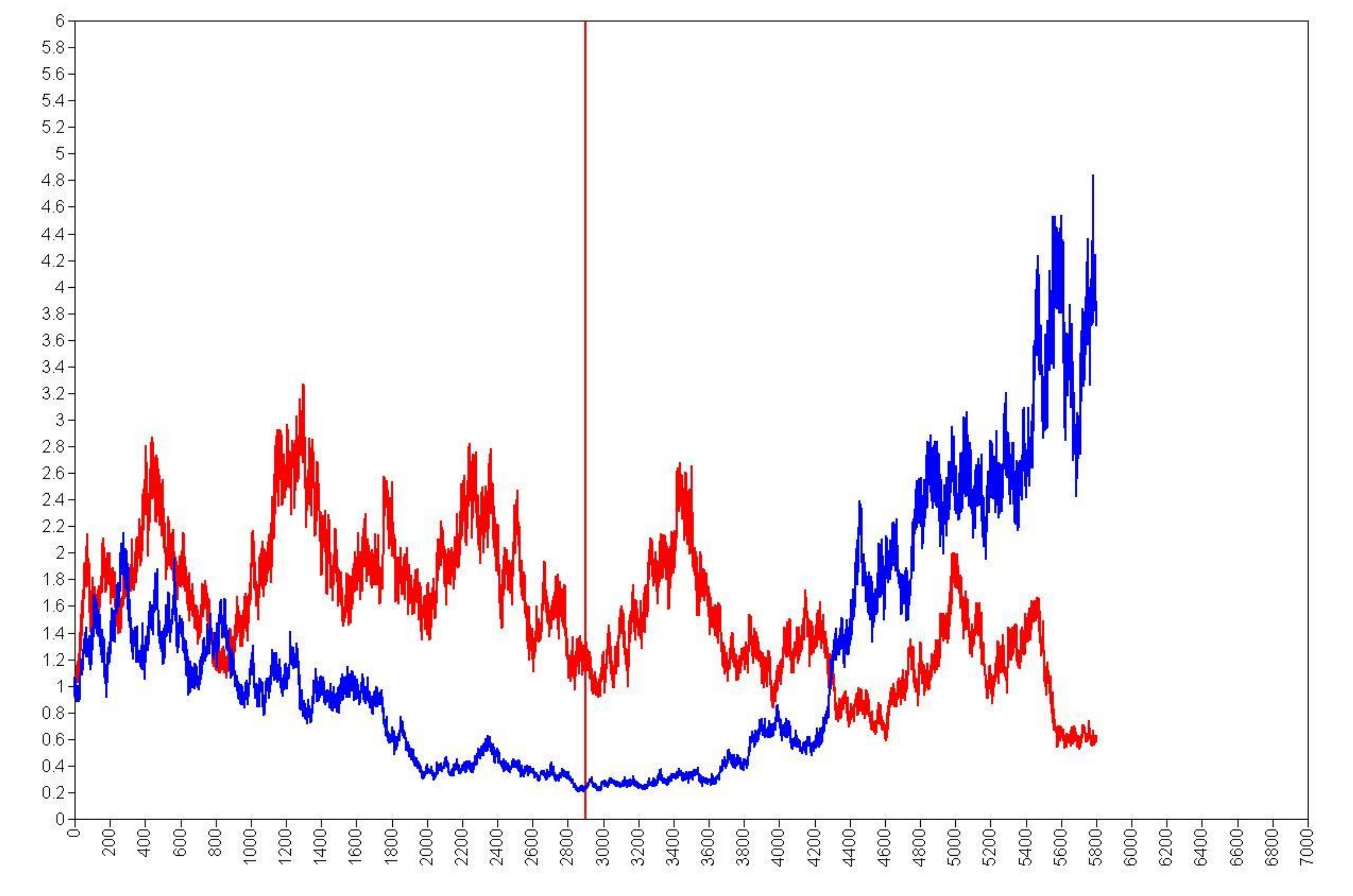


Convergence????

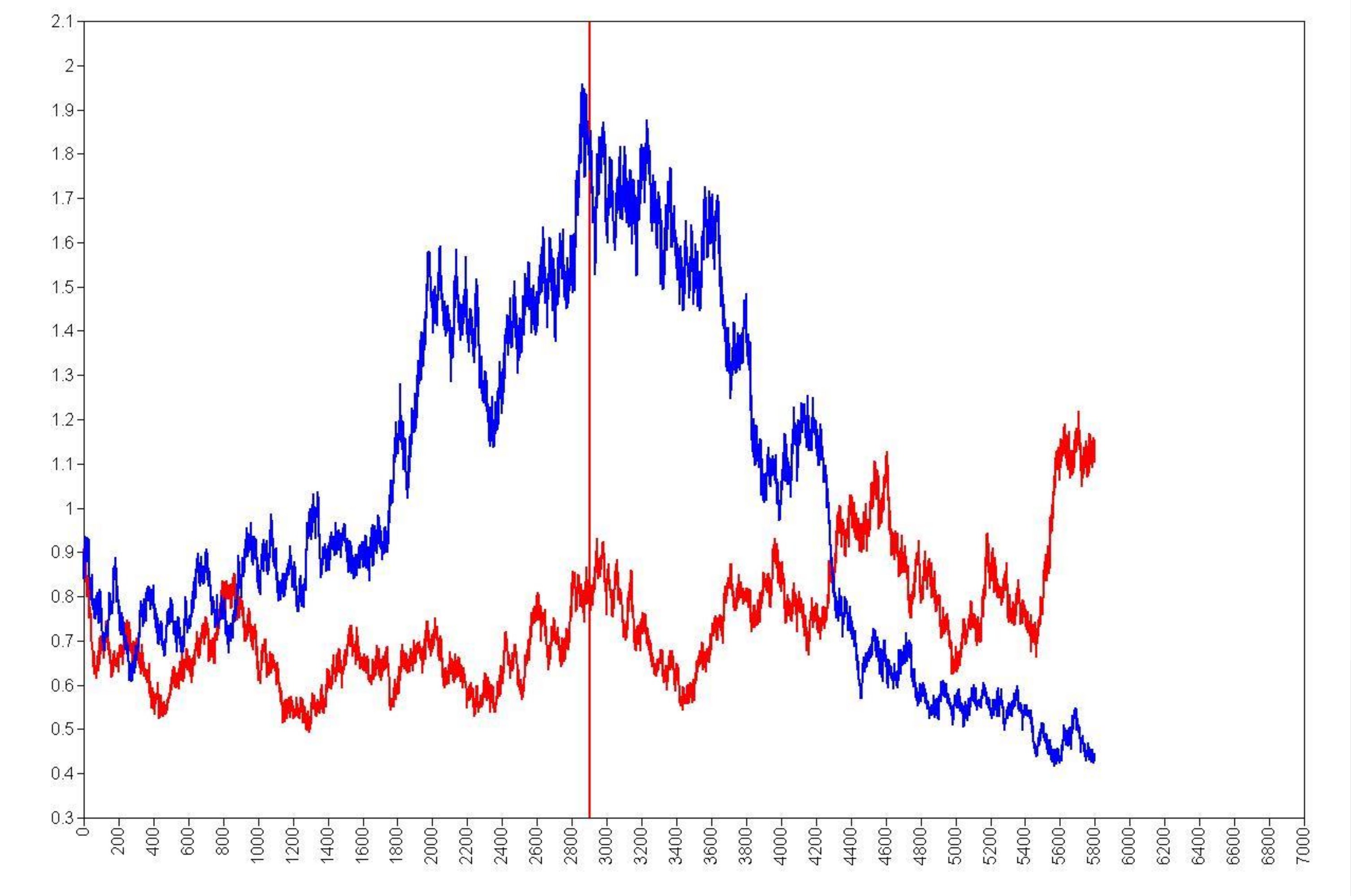


Universiteit Utrecht

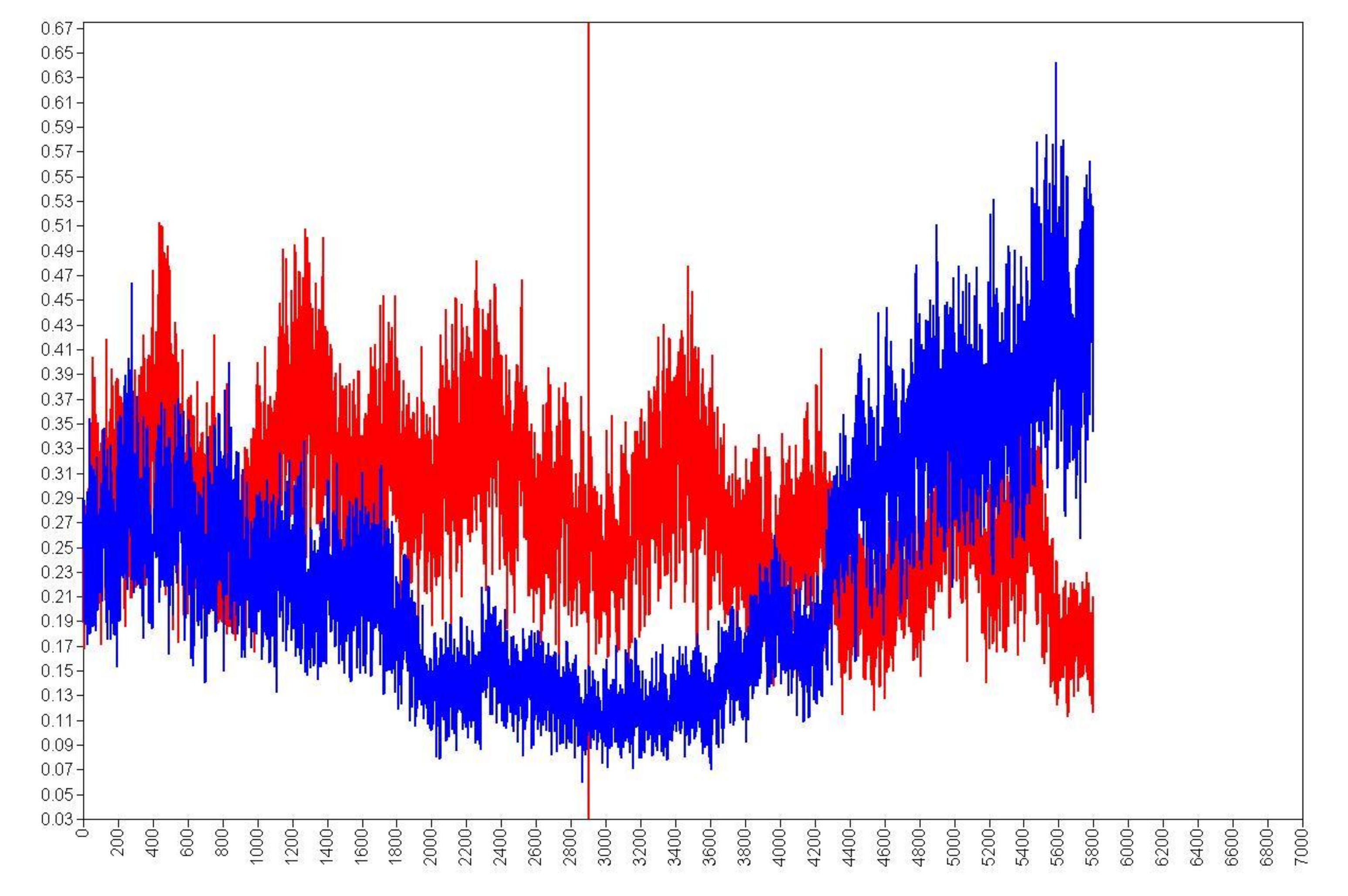
Convergence



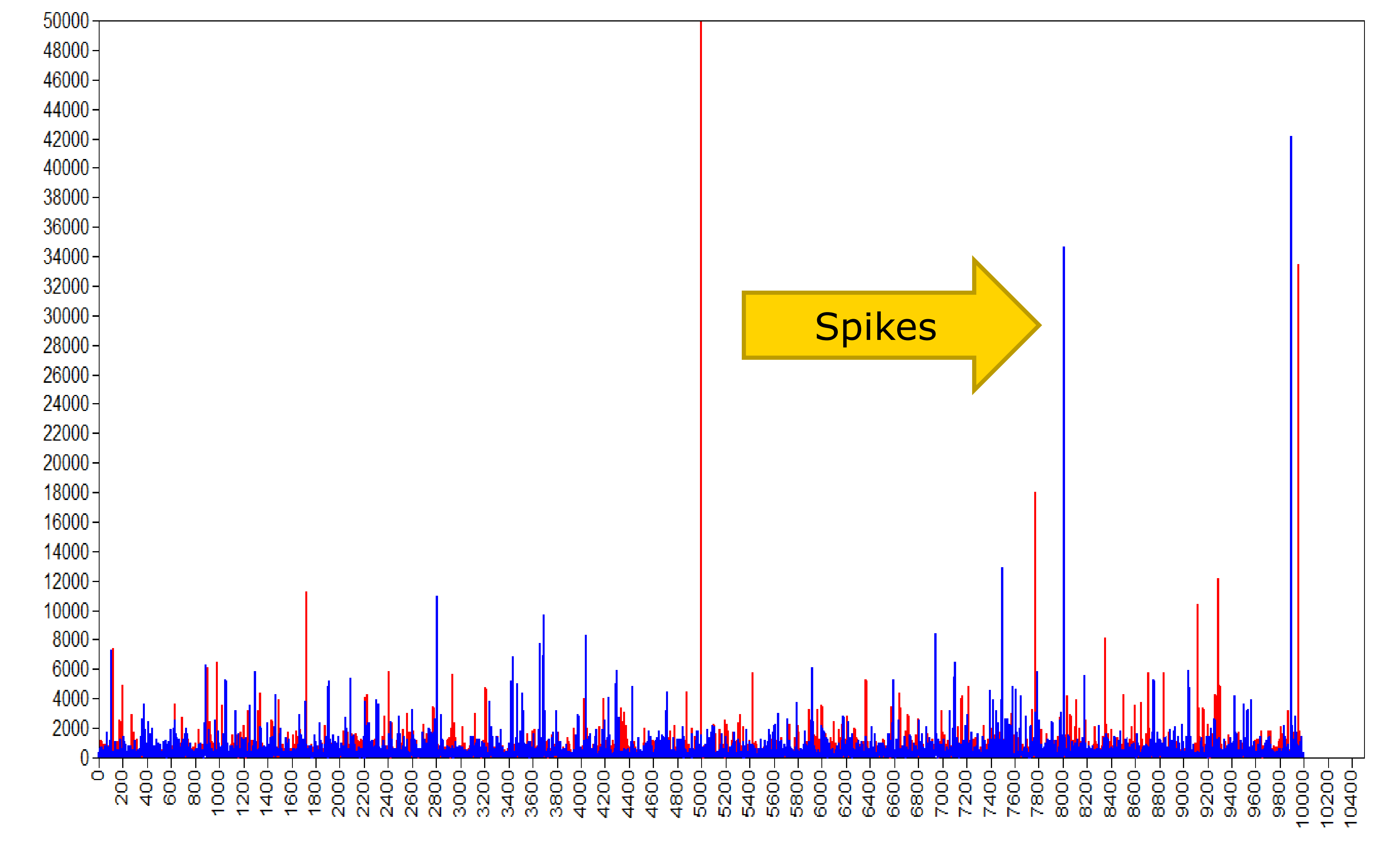
Convergence



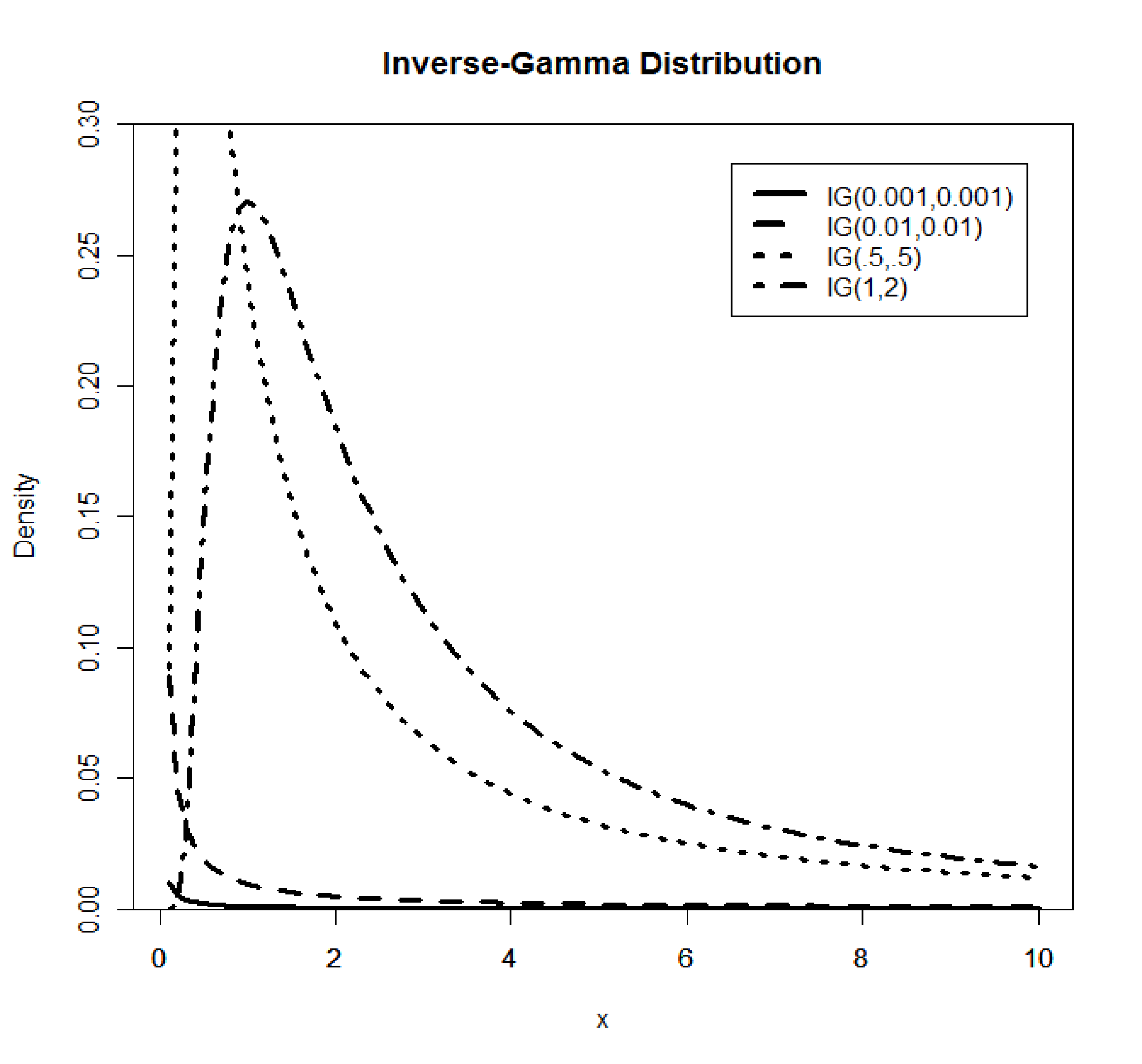
Convergence



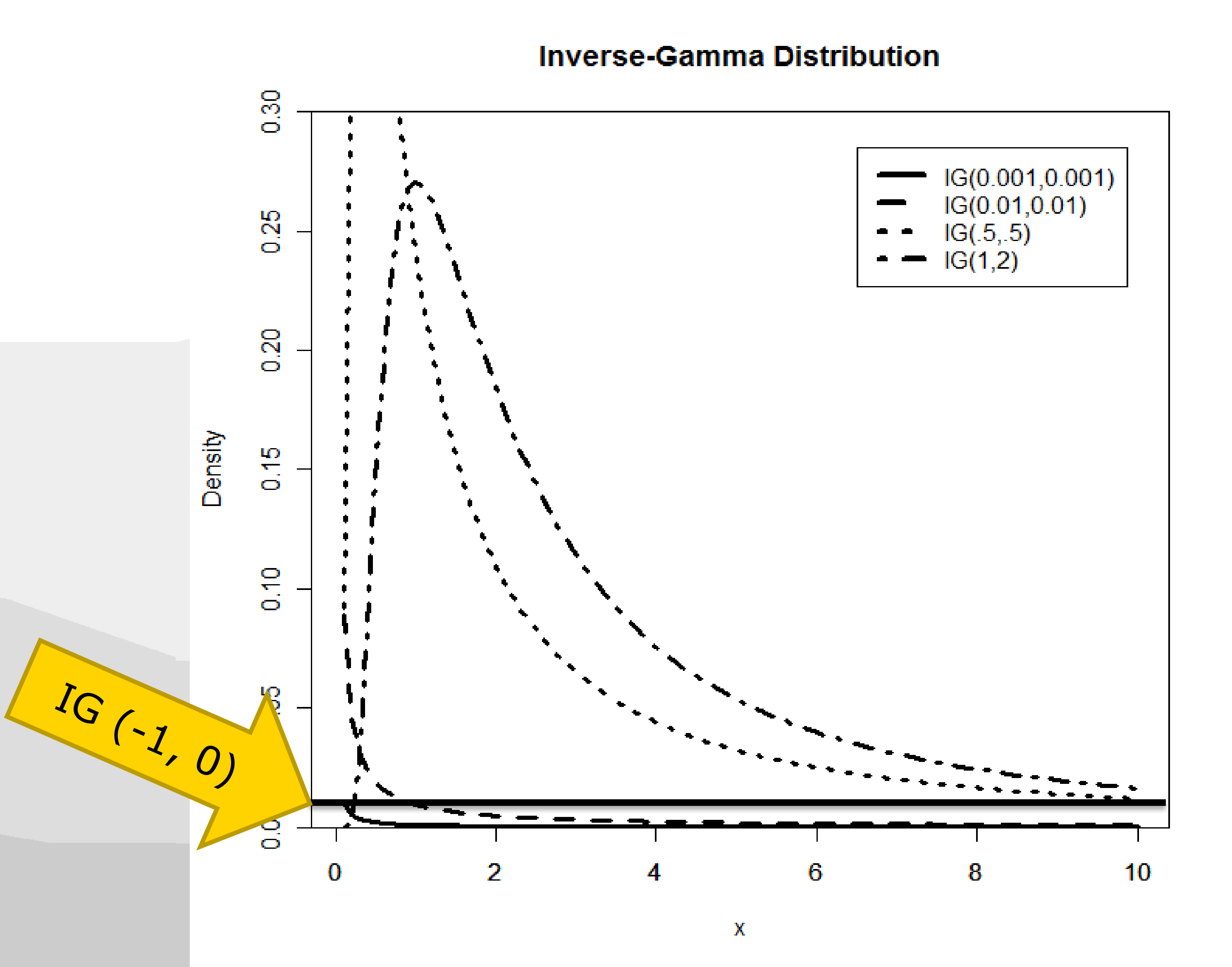
Trace plot for the variance of the Slope
Default prior setting IG(-1,0)



Default Prior settings



Default Prior settings



Universiteit Utrecht

Improper prior

- Probability distribution does not sum or integrate to one
- Shape and scale parameter need to be larger than zero

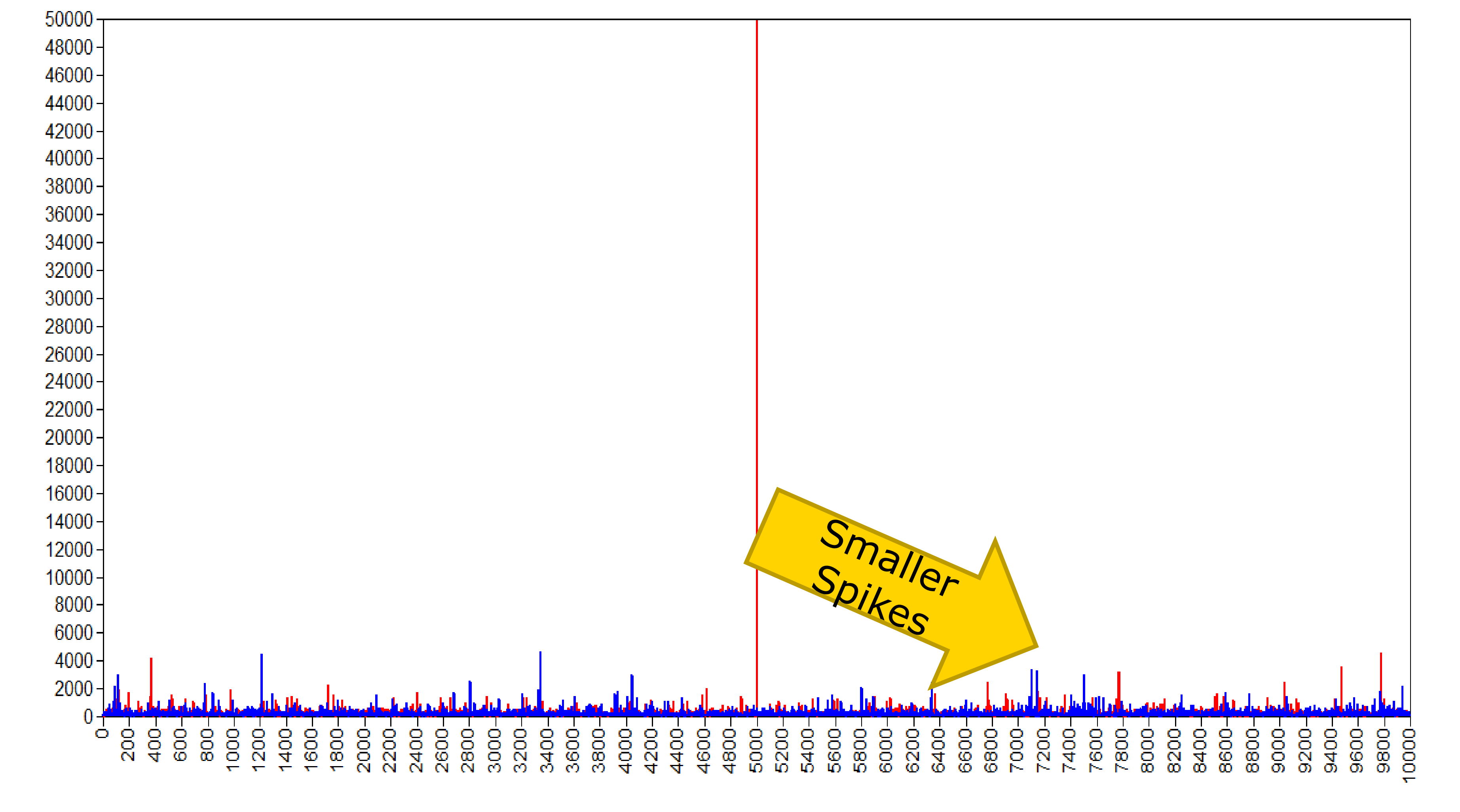
Improper prior:

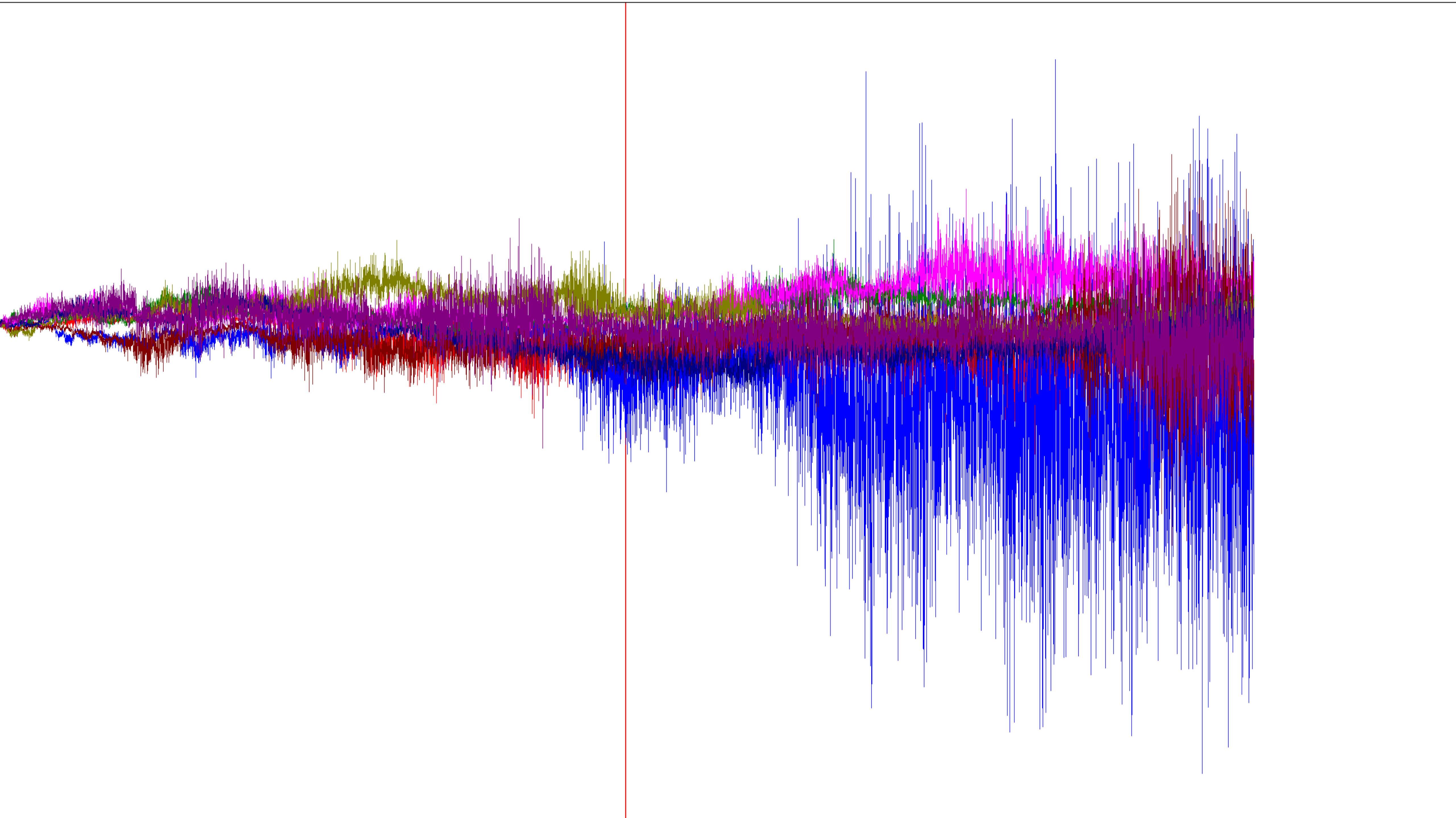
$p(\theta_l) \sim \mathrm{IG}(-1,0)$
$p(\theta_l) \sim \mathrm{IG}(0,0)$

Proper prior:

$p(\theta_l) \sim \mathrm{IG}(.001,.001)$
$p(\theta_l) \sim \mathrm{IG}(.5,.5)$

Trace plot for the variance of the Slope Prior setting IG(0,0)





Why Bayes?

Universiteit Utrecht

Why do researchers use Bayes

- First: Because they like the Bayesian paradigm!

Why do researchers use Bayes

- First: Because they like the Bayesian paradigm!
- Key difference between Bayesian statistical inference and conventional, frequentist estimation concerns the nature of the unknown parameters in the model

Why do researchers use Bayes

- In **frequentist framework** it is assumed that in the population there is only one true population parameter, for example, one true regression coefficient.

Why do researchers use Bayes

- In the **Bayesian framework**, there are two main ways of viewing the population parameter:
  - First, in the Bayesian view of subjective probability, all unknown parameters can be treated as unknown and fixed (Gelman & Robert, 2013). Bayesians can then model these unknown parameters as being random through a prior probability distribution (or *prior*) that captures the user's uncertainty of the fixed value of the parameter.

Why do researchers use Bayes

- In the **Bayesian framework**, there are two main ways of viewing the population parameter:

  - The second way of viewing parameters in the Bayesian framework is to recognize that the parameter of interest behaves in a stochastic fashion (i.e., it is not a fixed value but rather is random with an unknown probability distribution). In turn, the population parameter can be represented by a probability distribution with an unknown mean and variance since the parameter value is viewed as random under this definition, and this distribution is specified in part through the prior (Gill, 2008).

Universiteit Utrecht

Why do researchers use Bayes

- Interpretation confidence interval / credibility interval

Why do researchers use Bayes

What does *95% confidence interval* actually mean?

Why do researchers use Bayes

What does *95% confidence interval* NOT mean?

- We have a 95% probability that the true population value θ is within the limits of our confidence interval

Why do researchers use Bayes

What does *95% confidence interval* NOT mean?

- ~~We have a 95% probability that the true population value θ is within the limits of our confidence interval~~

Why do researchers use Bayes

What does *95% confidence interval* NOT mean?

- ~~We have a 95% probability that the true population value θ is within the limits of our confidence interval~~

- We only have an aggregate assurance that in the long run 95% of our confidence intervals contain the true population value

Universiteit Utrecht

Why do researchers use Bayes

What does a *95% central credibility interval* mean?

- We have a 95% probability that the population value θ is within the limits of our confidence interval

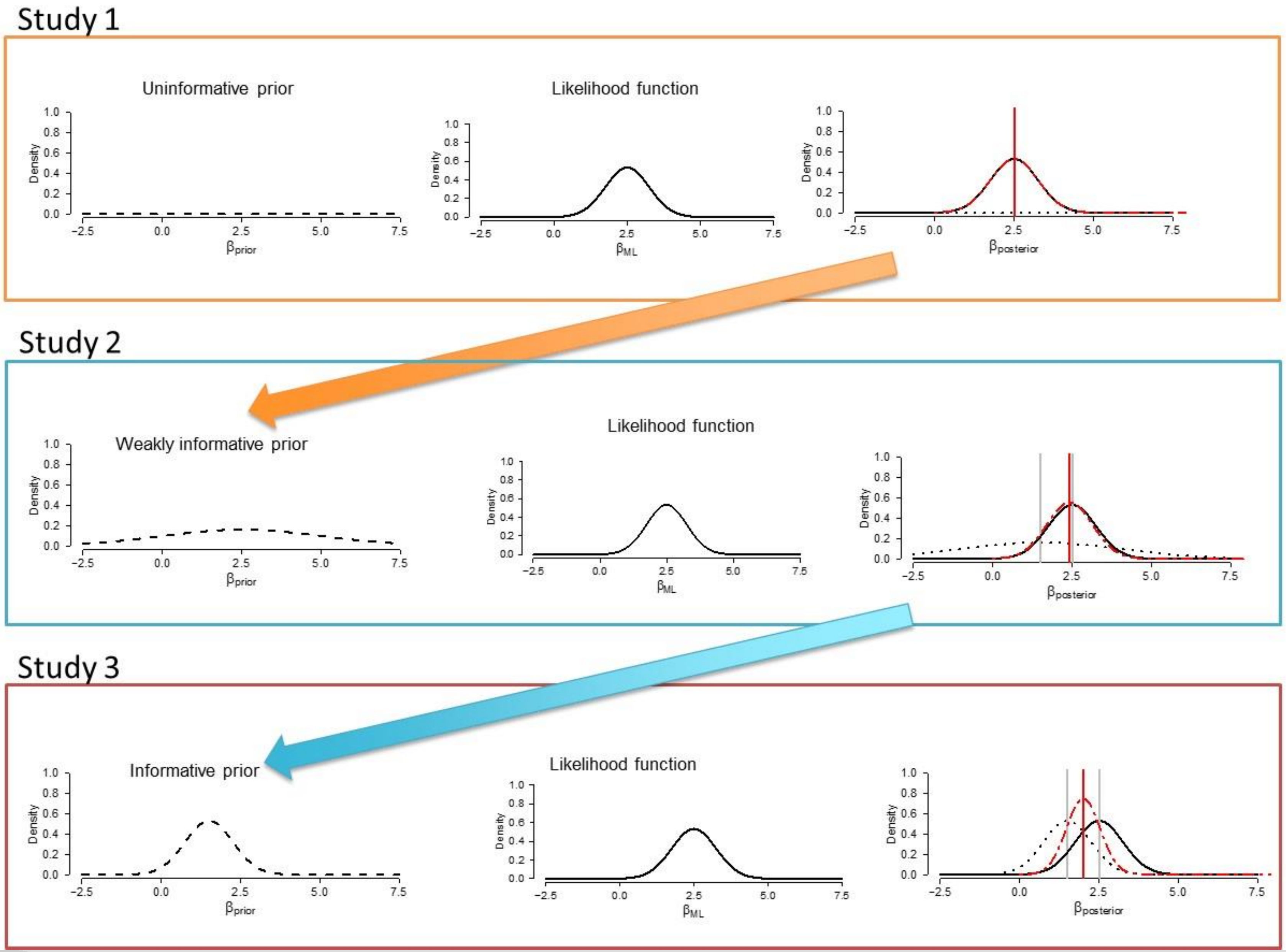
Why do researchers use Bayes

- Technical reasons:
  - complex models simply cannot be estimated using conventional statistics
  - to improve convergence issues
  - aid in model identification
  - produce more accurate parameter estimates.

Why do researchers use Bayes

- incorporate (un)certainty about a parameter and update this knowledge through the prior distribution.

Universiteit Utrecht



van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf , J.B., Neyer, F.J. & van Aken, M.A.G. (2014). A Gentle Introduction to Bayesian Analysis: Applications to Research in Child Development. Child Development, 85 (3), 842–860.

Why do researchers use Bayes

- Bayes is not based on large samples (i.e., the central limit theorem) and hence large samples are not required to obtain accurate results.

How large should the sample size be at the highest

level in multilevel analyses

????

With ML-estimation:

-> Boomsma (1983): 200 OK, at least 100

-> Hox, Maas Brinkhuis (2010): at least 100 groups

Universiteit Utrecht

With ML-estimation:

-> Boomsma (1983): 200 OK, at least 100
-> Hox, Maas Brinkhuis (2010): at least 100 groups

With Bayesian estimation:

-> Hox et al (2012): 20-25 OK!

Hox, J., van de Schoot. R., & Matthijsse, S. (2012). How few countries will do? Comparative survey analysis from a Bayesian perspective. Survey Research Methods, 6, 87-93.

Why do researchers use Bayes

- Even more reasons:
  - Non-normal data
  - Computational power
  - Missing data handling
  - Flexibiltiy
  - …

Exercise 2

OPEN ACCESS Freely available online PLOS | ONE

What Took Them So Long? Explaining PhD Delays among Doctoral Candidates

Rens van de Schoot[1,2]*, Mara A. Yerkes[3,4], Jolien M. Mouw[5], Hans Sonneveld[6,7]

1 Department of Methods and Statistics, Utrecht University, Utrecht, The Netherlands, 2 Optentia Research Focus Area, North-West University, Vanderbijlpark, South Africa, 3 Institute for Social Science Research, University of Queensland, Brisbane, Australia, 4 Erasmus University Rotterdam, Rotterdam, The Netherlands, 5 Education and Child Studies, Faculty of Social and Behavioural Sciences, Leiden University, Leiden, The Netherlands, 6 Netherlands Centre for Graduate and Research Schools, Utrecht, The Netherlands, 7 Tilburg Law School, Tilburg University, Tilburg, The Netherlands

Abstract

A delay in PhD completion, while likely undesirable for PhD candidates, can also be detrimental to universities if and when PhD delay leads to attrition/termination. Termination of the PhD trajectory can lead to individual stress, a loss of valuable time and resources invested in the candidate and can also mean a loss of competitive advantage. Using data from two studies of doctoral candidates in the Netherlands, we take a closer look at PhD duration and delay in doctoral completion. Specifically, we address the question: Is it possible to predict which PhD candidates will experience delays in the completion of their doctorate degree? If so, it might be possible to take steps to shorten or even prevent delay, thereby helping to enhance university competitiveness. Moreover, we discuss practical do's and don'ts for universities and graduate schools to minimize delays.

Citation: van de Schoot R, Yerkes MA, Mouw JM, Sonneveld H (2013) What Took Them So Long? Explaining PhD Delays among Doctoral Candidates. PLoS ONE 8(7): e68839. doi:10.1371/journal.pone.0068839

- 333 PhD recipients in The Netherlands
- how long it had taken them to finish their PhD thesis

  => 59.8 months
- difference between planned and actual project time in months

  => $M = 9.97, min\ /\ max = -31/91, SD = 14.43$
- assume we are interested in the question whether age (M=31.68, min/max=26/69) of the PhD recipients is related to delay in their project.
- assume we expect this relation to be non-linear.

pps-App - Rens van de Schoot

https://www.rensvandeschoot.com/tutorials/pps-app/

Rens van de /SCHOOT

Home About Rens Projects Publications Workshops Colleagues Rens's Ramblings Media Contact

pps-App

Home > Online Stats Training > pps-App

Plausible Parameter Space!

- Introduction
- Step 1. Set up Parameter Space
- Step 2: Set prior regression coefficients
- Step 3: Quantify uncertainty
- Step 4: Your Priors

Utrecht University

Influence of Priors

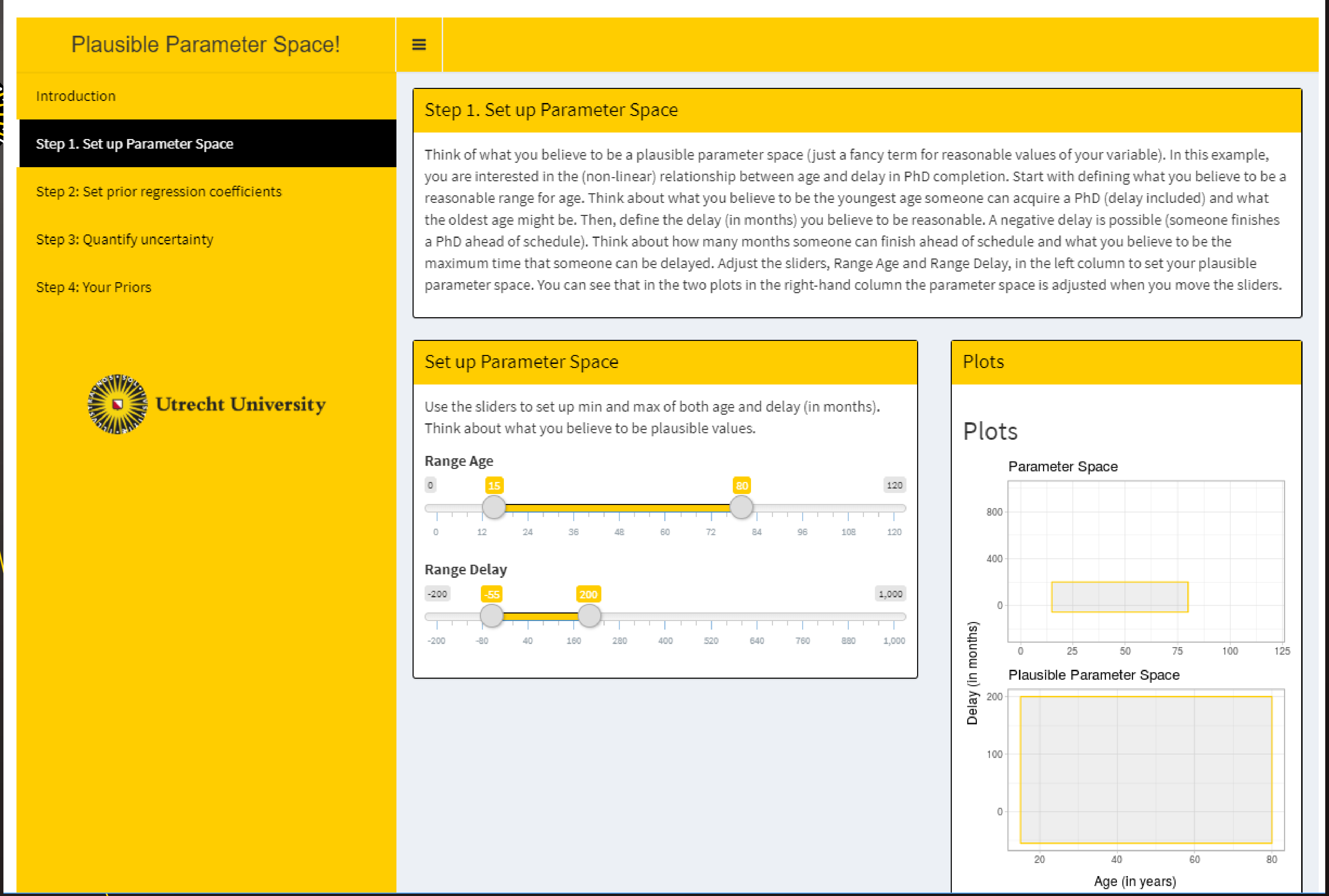
Version 0.3.2, created by Laurent Smeets and Rens van de Schoot

Show Disclaimer

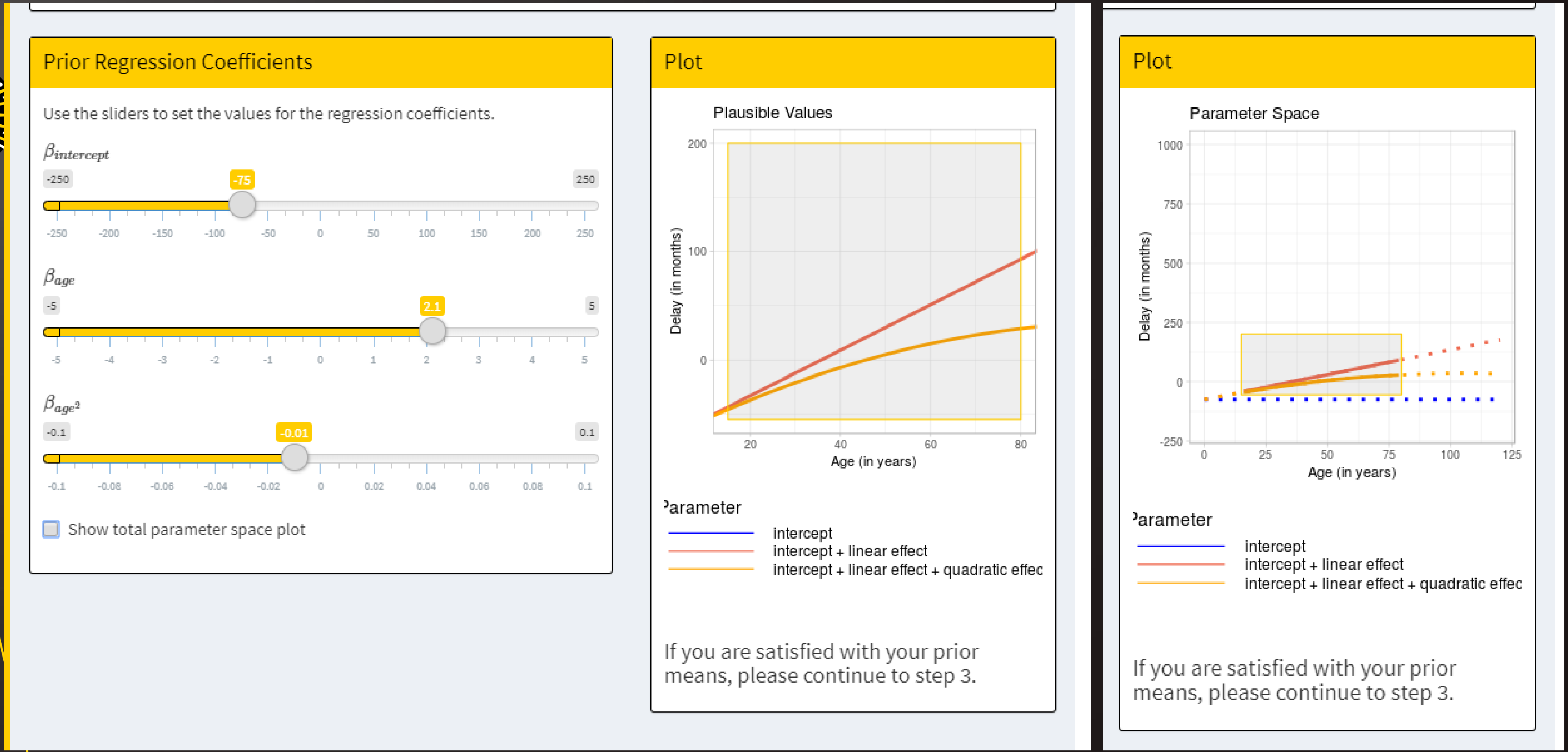
This Shiny App is designed to help users define their priors in a linear regression with two regression coefficients. Using the same example as in the software tutorials on this website, users are asked to specify their plausible parameter space and their expected prior means and uncertainty around these means. The Ph.D.-delay example has been used an easy-to-go introduction to Bayesian inference. In this example the linear and quadratic effect of age on Ph.D.-delay are estimated. Users learn about the interaction between a linear and a quadratic effect in the same model, about how to think about plausible parameter spaces, and about specification of normally distributed priors for regression coefficients.

The data is based on data described in Van de Schoot, R., Yerkes, M.A., Mouw, J.M. & Sonneveld, H. (2013). What Took Them So Long? Explaining PhD Delays among Doctoral Candidates. PLoS One, 8(7): e68839.

www.rensvandeschoot.com/pps-App



www.rensvandeschoot.com/pps-App



www.rensvandeschoot.com/pps-App

Standard Deviations Prior Regression Coefficients

Use the sliders to set the values for the prior variances (expressed in sd) of the regression coefficients.

Standard Deviation $\beta_{intercept}$

0 — 151 — 1,000

Standard Deviation β_{age}

0 — 14 — 100

Standard Deviation β_{age^2}

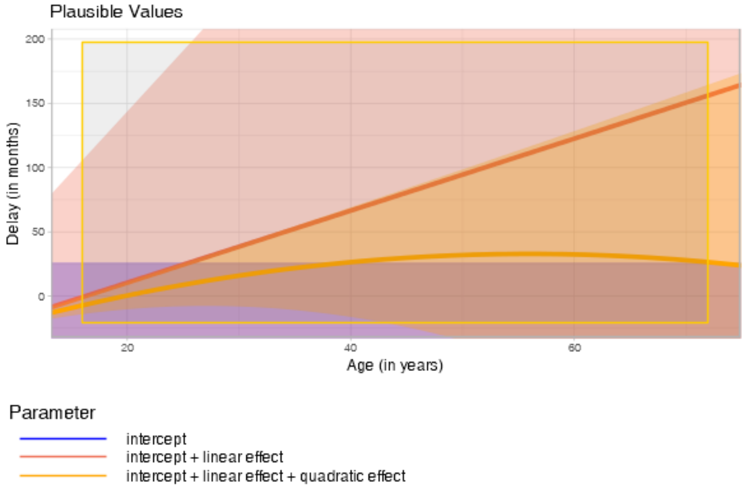
0 — 0.056 — 1

Percentage of uncertainty sd ribbons represent

66.7%

- [ ] Join Variances
- [ ] Show total parameter space plot

Plot



If you are satisfied with your priors, please have a look at them in the next tab.

Your priors

Different software requires different specification of the hyperparameters. Look at the specification that is relevant for you.

Parametrized as N(mean, variance)

The hyperparameters of the priors you have selected correspond to:

- Intercept ($\beta_{intercept}$) ~ N(-75, 71)
- regression coefficient age (β_{age}) ~ N(2.1, 26)
- regression coefficient squared (β_{age^2}) ~ N(-0.01, 0.263)

Parametrized as N(mean, sd)

The hyperparameters of the priors you have selected correspond to:

- Intercept ($\beta_{intercept}$) ~ N(-75, 8.4)
- regression coefficient age (β_{age}) ~ N(2.1, 5.1)
- regression coefficient squared (β_{age^2}) ~ N(-0.01, 0.51)

Parametrized as N(mean, precision), precision is 1/variance.

The hyperparameters of the priors you have selected correspond to:

- Intercept ($\beta_{intercept}$) ~ N(-75, 0.014)
- regression coefficient age (β_{age}) ~ N(2.1, 0.038)
- regression coefficient age squared(β_{age^2}) ~ N(-0.01, 3.8)

www.rensvandeschoot.com/pps-App

Your priors

Different software requires different specification of the hyperparameters. Look at the specification that is relevant for you.

Parametrized as N(mean, variance)

The hyperparameters of the priors you have selected correspond to:

- Intercept ($\beta_{intercept}$) ~ N(-75, 71)
- regression coefficient age (β_{age}) ~ N(2.1, 26)
- regression coefficient squared (β_{age^2}) ~ N(-0.01, 0.263)

Parametrized as N(mean, sd)

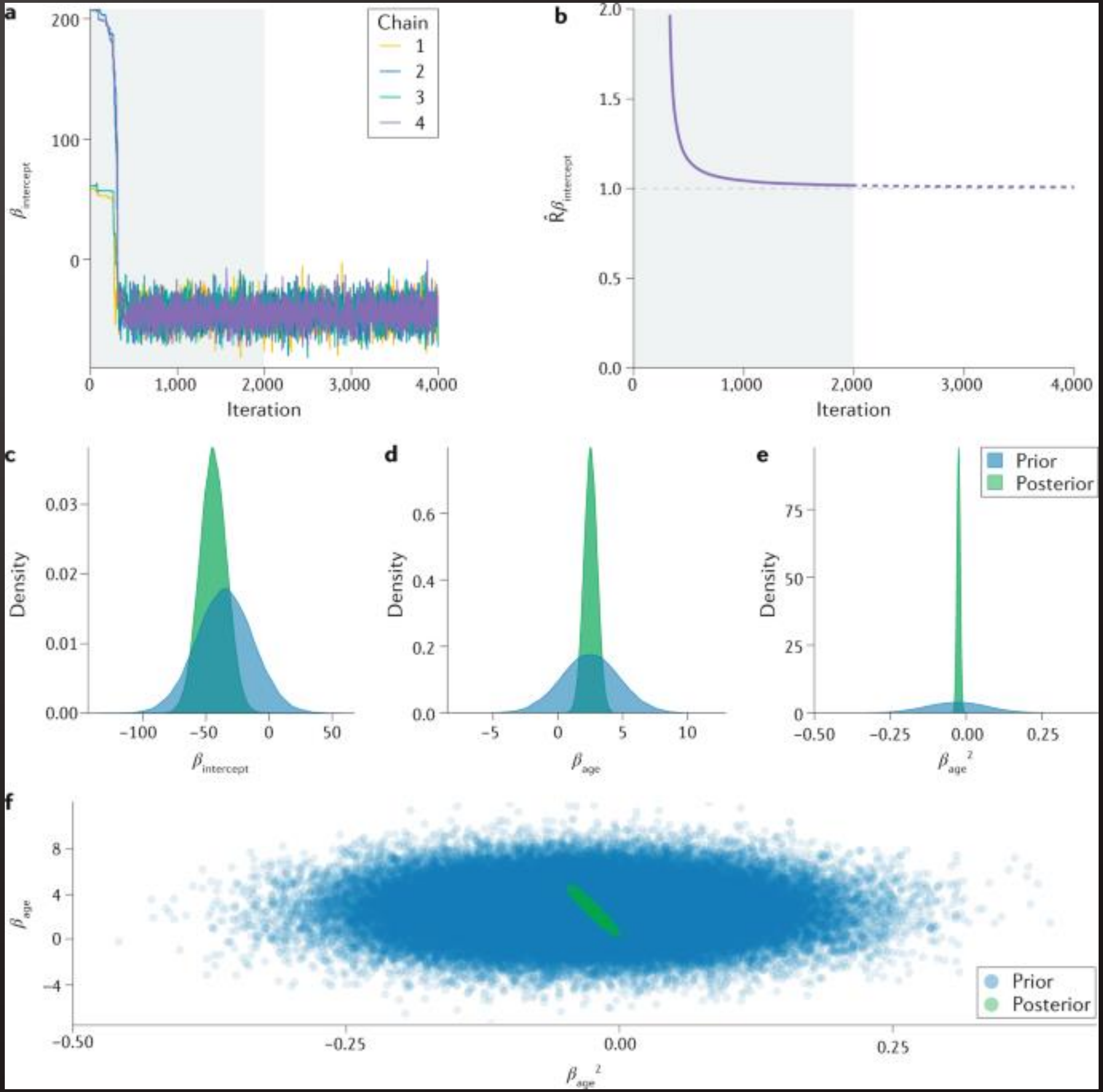
The hyperparameters of the priors you have selected correspond to:

- Intercept ($\beta_{intercept}$) ~ N(-75, 8.4)
- regression coefficient age (β_{age}) ~ N(2.1, 5.1)
- regression coefficient squared (β_{age^2}) ~ N(-0.01, 0.51)

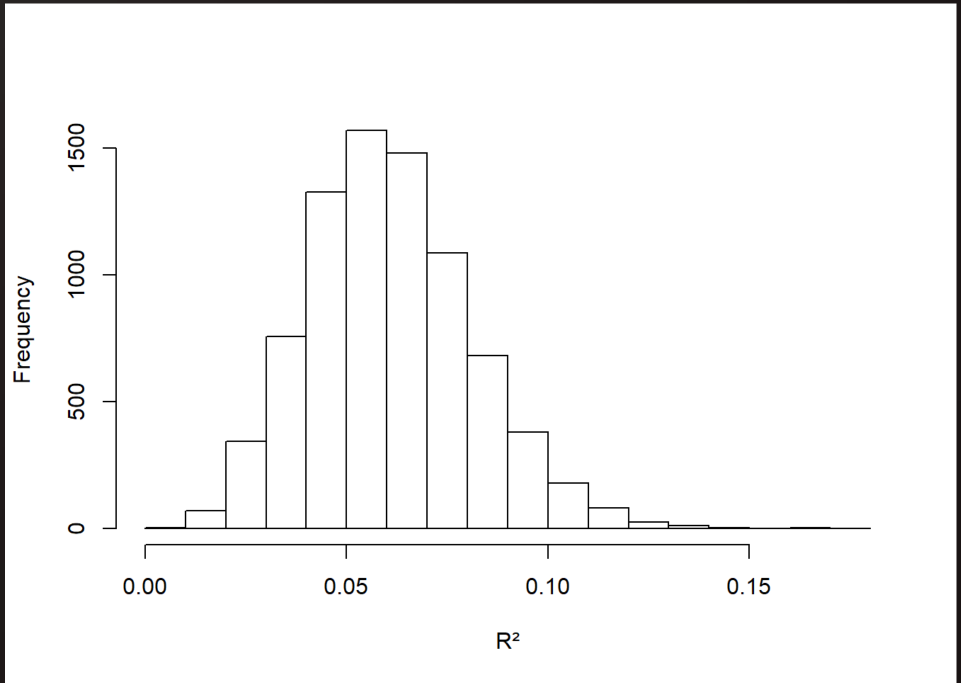
Parametrized as N(mean, precision), precision is 1/variance.

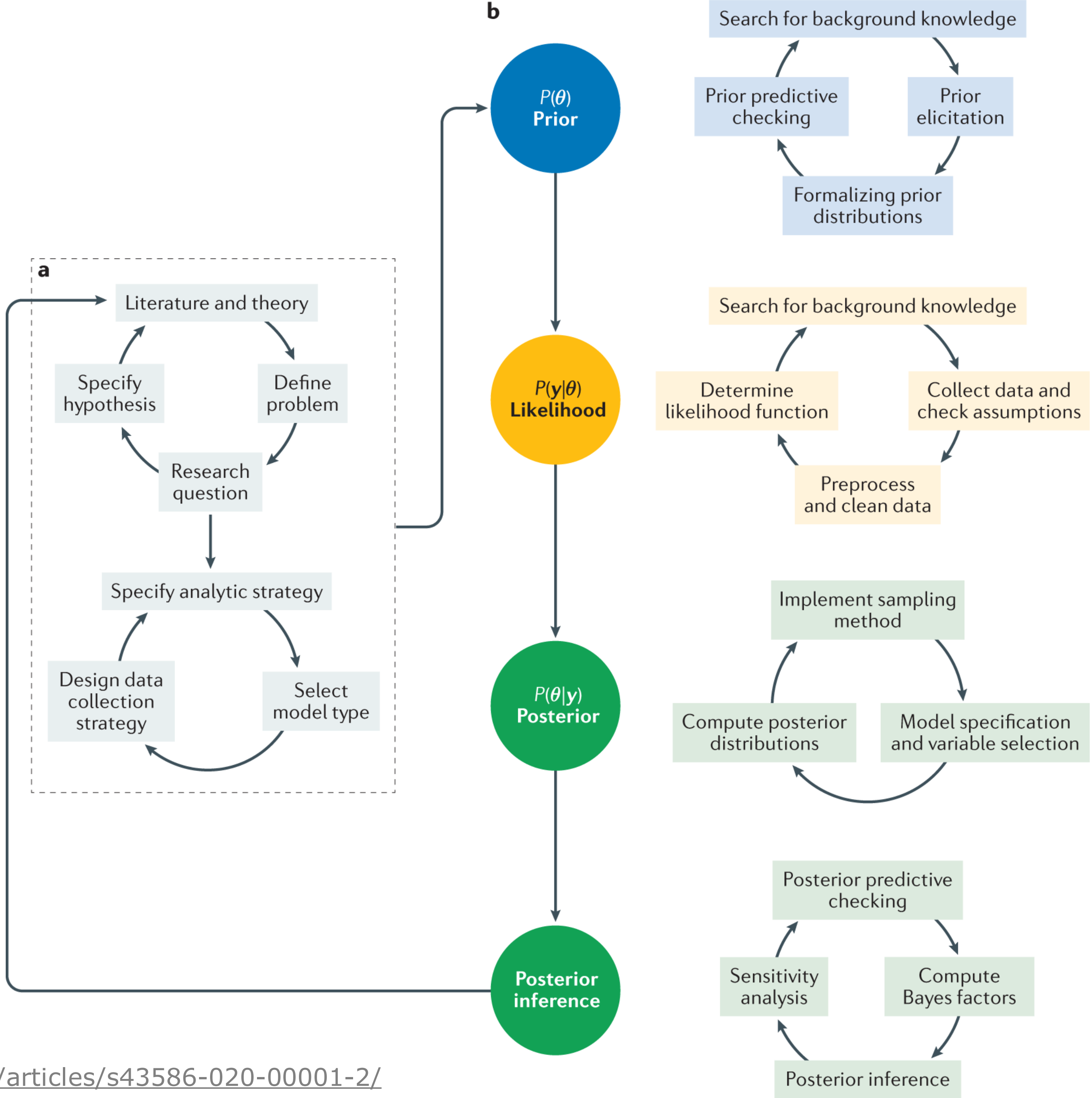
The hyperparameters of the priors you have selected correspond to:

- Intercept ($\beta_{intercept}$) ~ N(-75, 0.014)
- regression coefficient age (β_{age}) ~ N(2.1, 0.038)
- regression coefficient age squared(β_{age^2}) ~ N(-0.01, 3.8)


a
Chain
1
2
3
4
$\beta_{intercept}$
Iteration
b
$\hat{R}\beta_{intercept}$
Iteration
c
d
e
Density
Prior
Posterior
$\beta_{intercept}$
β_{age}
β_{age^2}
f
β_{age}
β_{age^2}
Prior
Posterior

	mean	sd	2.5%	50%	97.5%
beta_intercept	-44.425	10.579	-64.325	-44.668	-23.387
beta_age	2.532	0.503	1.522	2.544	3.477
beta_age2	-0.025	0.005	-0.034	-0.025	-0.014
epsilon2	196.923	15.266	168.758	196.255	228.166





www.nature.com/articles/s43586-020-00001-2/