Bayesian Inference Project

Carlos Enmanuel Soto López[1,2]
carlos.soto362@gmail.com

`Supervisors: Paolo Lazzari`[2] `and Fabio Anselmi`[1,2].

1. Università degli Studi di Trieste, 31127, Italy
2. Istituto nazionale di oceanografia e di geofisica sperimentale - OGS, Trieste, 34010, Italy

Overview

1. **Bayesian experiment design**
2. **Problem 1**
   2.1 Description
3. **Problem 2**
   3.1 Description
4. **Problem 3**
   4.1 Description

Statement of the problem

- Data collected d $= \{(x_1, y_1)(x_2, y_2)\}$.
- The posterior of these data is the prior for a future experiment.
- We want to choose the best future measurement x_3 such that it maximises the expectation of a utility funciton.

$$\mathbb{E}[d|e] = \int d\theta U(\theta, d, e) Pr(\theta|d) \tag{1}$$

Bayesian experiment design

These problem is in the category of Bayesian Experiment design,

current Goal

A design η has to be chosen from a set $\mathcal{H}$ (e.g. the fraction η_i of people taking the drug i).

The expected experiment

Given the design, we will have an output y. A stochastic function, with parameters θ, maps the design to the outputs y.

How to pursue the goal

We want to optimize the design with respect to the expected value of a utility function U (our expected gain in information from the experiment).

Description of problem 1

Consider the linear model,

$$y = \theta x + \epsilon \tag{2}$$

Adopt as utility function the inverse of the posterior variance on θ *from both the past experiment and e, and assuming constant noise,* $\tau(x) = \tau$*, determine what is the best choice of a single future measurement,* x_f *, in the range* $[0, x_{max}]$*:*

Choose a prior

I used a Jeffrey's prior,

$$
\begin{aligned}
p(\theta) &= \sqrt{-\mathbb{E}\left[\left(\frac{d}{d\theta^2}\log p(y_1|\theta;\sigma)\right)\right]} \\
&= \sqrt{-\mathbb{E}\left[\left(\frac{d}{d\theta^2}\frac{-(\theta x_1 - y_1)^2}{2\sigma^2}\right)\right]} \\
&= \left|\frac{x_1}{\sigma}\right|
\end{aligned}
\tag{3}
$$

Posterior of first experiment

I did the computations, but there is an easy rule called precision-weighted update rule.

$$
\begin{aligned}
p(\theta|\{y_1, y_2\}; \sigma) &\propto \exp\left(\frac{-(\theta - \frac{y_1}{x_1})^2}{2\frac{\sigma^2}{x_1^2}}\right) \exp\left(\frac{-(\theta - \frac{y_2}{x_2})^2}{2\frac{\sigma^2}{x_2^2}}\right) \\
&\propto \exp\left(-\frac{(\theta - m_{1,2})^2}{2\sigma_{1,2}^2}\right), \qquad m_{1,2} = \frac{x_1 y_1 + x_2 y_2}{x_1^2 + x_2^2}, \qquad \sigma_{1,2}^2 = \frac{\sigma^2}{x_1^2 + x_2^2}.
\end{aligned} \tag{4}
$$

Posterior of second experiment

$$
\begin{aligned}
p(\theta|\{y_1, y_2, y_3\}; \sigma, \tau) &\propto \exp\left(-\frac{(\theta - m_{1,2})^2}{2\sigma_{1,2}^2}\right) \exp\left(-\frac{(\theta x_3 - y_3)^2}{2\tau^2}\right) \\
&\propto \exp\left(-\frac{(\theta - m_{3,\tau})^2}{2\sigma_{3,\tau}^2}\right), \\
m_{3,\tau} &= \frac{m_{1,2}\tau^2 + x_3 y_3 \sigma_{1,2}}{\tau^2 + \sigma_{1,2}^2 x_3^2}, \qquad \sigma_{3,\tau}^2 = \frac{\tau^2 + \sigma_{1,2}^2 x_3^2}{\sigma_{1,2}^2 \tau^2}.
\end{aligned}
\tag{5}
$$

Utility function

The utility function is the inverse of the variance,

$$U(x_3) = \sigma_{3,\tau}^{-2} = \left[\frac{x_1^2 + x_2^2}{\sigma^2} + \frac{x_3^2}{\tau^2} \right] \tag{6}$$

And is maximized with $x_f = x_max$.

Description of problem 2

Consider now the case where we could build a more accurate experiment, a, which could measure y with noise $\tau^* << \tau$, but only if the dependent variable falls around a certain value, y^*. We can model this by writing for the noise of a as:

$$\tau_a^2 = \tau^* exp\left(\frac{(y - y^*)^2}{2\Delta^2}\right) \tag{7}$$

We consider the parameters y^* , τ^* , Δ as fixed quantities. Find the optimal location x_f for a new measurement and compare the performance of a with the experiment considered above, e. Discuss how and why the result depends on the ratio $\Delta/\Delta y$, where Δy is the present-day uncertainty in the value of y at the location x_f. Produce numerical results for the following choices: $x_0 = 0.5$, $x_1 = 1.0$, $x_{max} = 3$, $\sigma = 0.1$, $\theta = 1.0$, $\tau = 0.1$, $\tau^* = \sigma/5$, $y^* = 1.5$, $\Delta = 0.1$. Plot the expected utility as a function of Δ (all other quantities constant) and determine the cross-over value for Δ at which the experimental configuration to be preferred changes from e to a.

Expected Utility function

$$
\begin{aligned}
\mathbb{E}[U(x_3,\theta)] &= \frac{x_1^2+x_2^2}{\sigma^2} + \frac{x_3^2}{\tau^*}\int_{-\infty}^{\infty} \exp\left(\frac{-(\theta x_3 - y^*)^2}{2\Delta^2}\right) p(\theta|\{y_1,y_2\},\sigma)d\theta \\
&= \frac{x_1^2+x_2^2}{\sigma^2} + \frac{x_3^2}{\sqrt{2\pi\tau^{*2}\sigma_{1,2}^2}}\int_{-\infty}^{\infty} \exp\left(\frac{-(\theta x_3 - y^*)^2}{2\Delta^2}\right) \exp\left(\frac{-(\theta - m_{1,2})^2}{2\sigma_{1,2}^2}\right) d\theta \\
&= \frac{x_1^2+x_2^2}{\sigma^2} + \frac{x_3^2}{\sqrt{\tau^{*2}(1+\frac{\sigma_{1,2}^2 x_3^2}{\Delta^2})}} \times \\
&\exp\left(-\frac{1}{2}\left[\left(\frac{m_{1,2}}{\sigma_{1,2}}\right)^2 + \left(\frac{y^*}{\Delta}\right)^2 - \frac{(\Delta^2 + x_3^2\sigma_{1,2}^2)(m_{1,2}\Delta^2 + x_3 y^* \sigma_{1,2}^2)}{\sigma_{1,2}^2\Delta^2}\right]\right)
\end{aligned} \tag{8}
$$

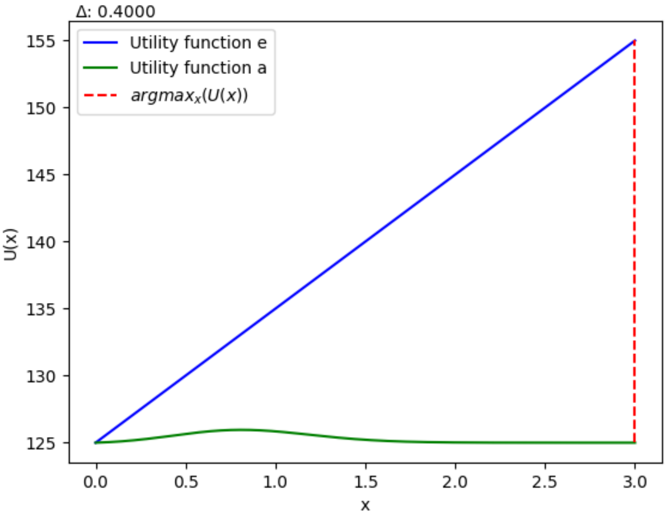
Maximum of the Utility function

We want to maximize this expression, as a function of x_3. After deriving, and equating to zero, the exponent and denominator factors cancel, and we end up with

$$\sigma_{1,2}^6 x_3^5 y^* + \sigma_{1,2}^4 \Delta^2 x_3^2 (x_3 y^* - 1) + m_{1,2}^2 \left(\sigma_{1,2}^2 \Delta^2 x_3^2 + \Delta^4\right) - m_{1,2} \left(\sigma_{1,2}^2 \Delta^4 x_3^2 + \Delta^6\right) = 0 \tag{9}$$

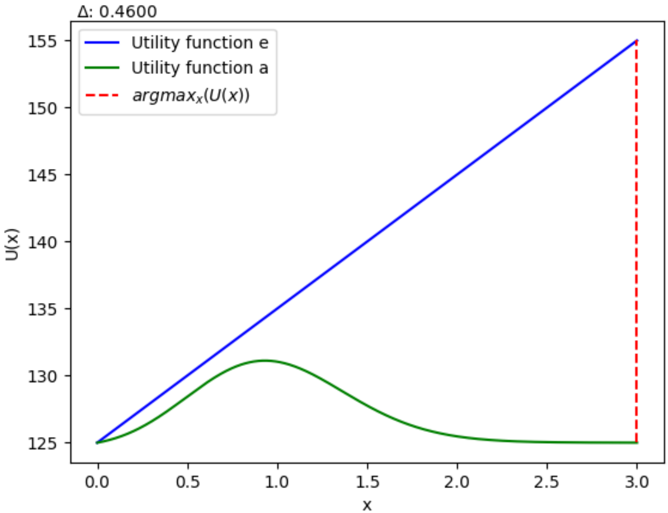
Not much to be done analytically! See here for the code use to analyze this problem numerically.

$\Delta = 0.4$



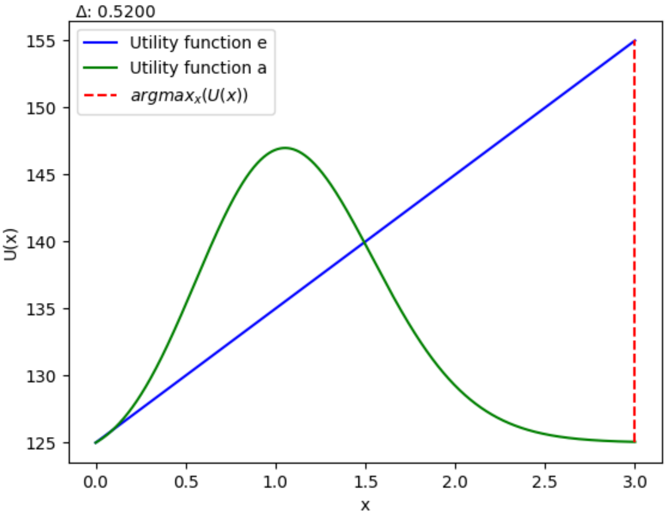
Δ: 0.4000
Utility function e
Utility function a
$argmax_x(U(x))$
U(x)
x
155
150
145
140
135
130
125
0.0
0.5
1.0
1.5
2.0
2.5
3.0

$\Delta = 0.46$



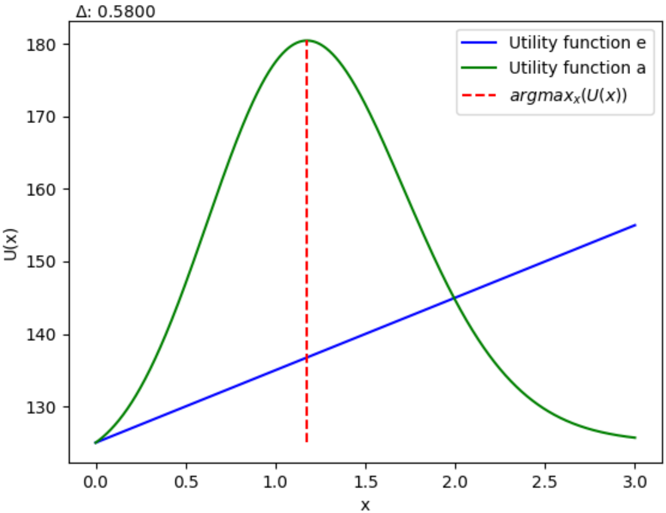
Δ: 0.4600
Utility function e
Utility function a
$argmax_x(U(x))$
155
150
145
140
135
130
125
U(x)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
x

$\Delta = 0.52$



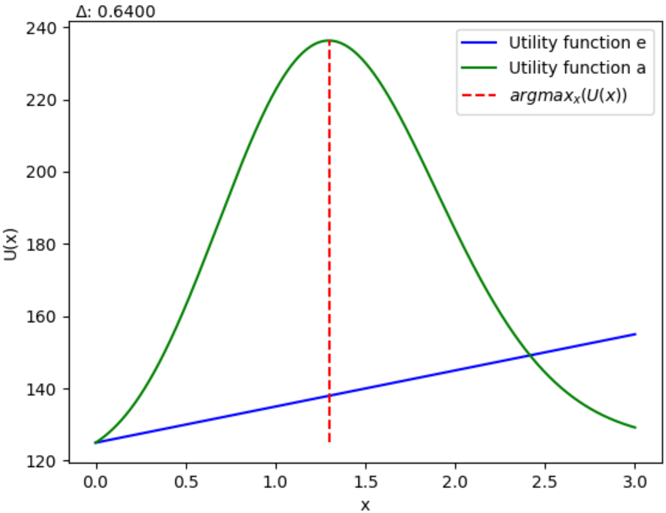
Δ: 0.5200
Utility function e
Utility function a
$argmax_x(U(x))$
155
150
145
140
135
130
125
U(x)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
x

$\Delta = 0.58$



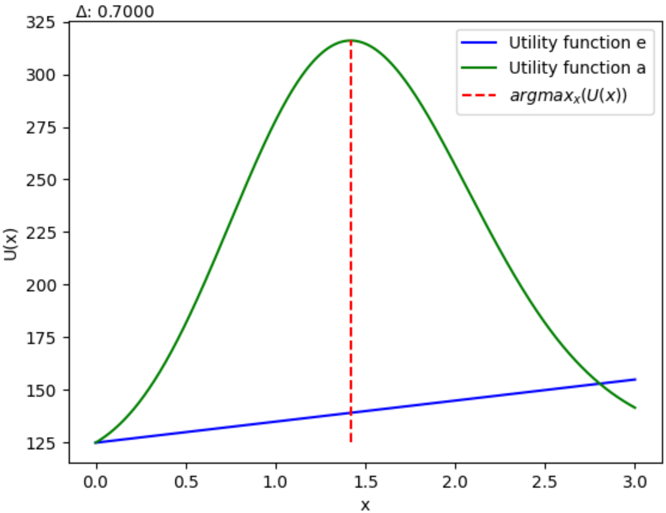
Δ: 0.5800
180
170
160
150
140
130
U(x)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
x
Utility function e
Utility function a
$argmax_x(U(x))$

$\Delta = 0.64$



Δ: 0.6400
Utility function e
Utility function a
$argmax_x(U(x))$
U(x)
x
240
220
200
180
160
140
120
0.0
0.5
1.0
1.5
2.0
2.5
3.0

$\Delta = 0.7$



Δ: 0.7000
325
300
275
250
225
200
175
150
125
U(x)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
x
Utility function e
Utility function a
$argmax_x(U(x))$

Description of problem 3

Consider now the task of optimizing the future experiment a to distinguish between two models for the data: Model 1 is the linear model above, Model 2 has an additional quadratic term and is given by:

$$y = \theta x + \psi x^2 + \epsilon \tag{10}$$

with ψ an additional parameter with a suitably chosen prior. As a utility function, adopt the volume of parameter space where the absolute value of the logarithm of the Bayes factor between the two models exceeds the threshold value of 2.5. This means that in such a region of parameter space the outcome of model comparison (using both past data and the future data point from a) would be above the 'moderate evidence' threshold (for either model) under the Jeffreys' scale. Determine the value of y^* (for fixed Δ) that maximizes such a utility function and discuss the result by comparing with maximization of the previous utility function.

Posterior probability

$$p(\theta|\{y_1, y_2, y_3\}, \psi; \sigma, \tau) \propto \exp\left(-\frac{(\theta - m_{3,\tau,\psi})^2}{2\sigma^2_{3,\tau,\psi}}\right),$$

$$m_{3,\tau,\psi} = \frac{m_{1,2,\psi}\tau^2 + x_3\hat{y}_3\sigma_{1,2,\psi}}{\tau^2 + \sigma^2_{1,2,\psi}x_3^2}, \qquad \sigma^2_{3,\tau,\psi} = \frac{\tau^2 + \sigma^2_{1,2,\psi}x_3^2}{\sigma^2_{1,2,\psi}\tau^2}. \tag{11}$$

where $\hat{y} = y - \psi x^2$.

Bayes Factor (1)

$$p(d|M_1) = \frac{1}{(2\pi)\sigma\tau}\frac{1}{\sqrt{\sum_{i=1}^{3}\frac{x_i^2}{\sigma_i^2}}}\exp\left(-\frac{1}{2}\left[\sum_{i=1}^{3}\frac{y_i^2}{\sigma_i^2} - \frac{\left(\sum_{i=1}^{3}\frac{x_i y_i}{\sigma_i^2}\right)^2}{\sum_{i=1}^{3}\frac{x_i^2}{\sigma_i^2}}\right]\right) \tag{12}$$

Bayes Factor (2)

$$p(d|M_2) = \frac{1}{\sqrt{2\pi}\sigma\tau} \frac{1}{\sqrt{\sum_{i=1}^{3} \frac{x_i^2}{\sigma_i^2}}} \sqrt{\frac{1}{\sum_{i=1}^{3} \frac{x_i^4}{\sigma_i^2}}} \exp\left(-\frac{1}{2} \left[\sum_{i=1}^{3} \frac{y_i^2}{\sigma_i^2} - \frac{\left(\sum_{i=1}^{3} \frac{x_i y_i}{\sigma_i^2} \right)^2}{\sum_{i=1}^{3} \frac{x_i^2}{\sigma_i^2}} \right] \right) \quad (13)$$

Bayes Factor (3)

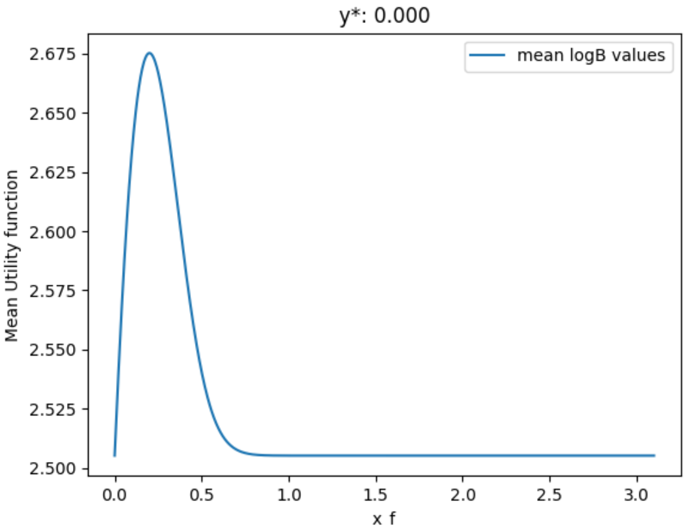
$$B = \frac{p(d|M_2)}{p(d|M_1)} = \frac{1}{\sqrt{\sum_{i=1}^{3} \frac{x_i^4}{\sigma_i^2}}} \tag{14}$$

$$logB = -\frac{1}{2}\log\left(\frac{x_1}{\sigma} + \frac{x_2}{\sigma} + \frac{x_3}{\tau}\right) \tag{15}$$

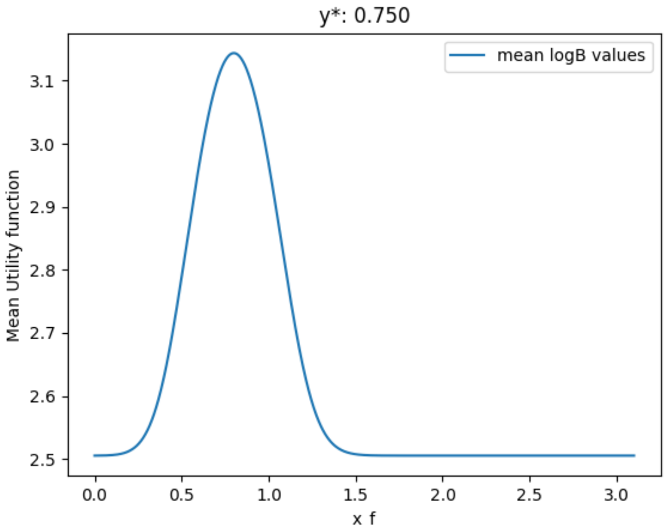
Some preliminary experiments

The problem talks about the volume of the parameters. First, I will work on this expression. I want the value of x_3 that maximizes the value of the logB if the data was generated with model 1, in such a way that I can optimally distinguish between the two models.
To test the computation, I made again a simulation, and compute many utility functions, but these time, changing y^*,

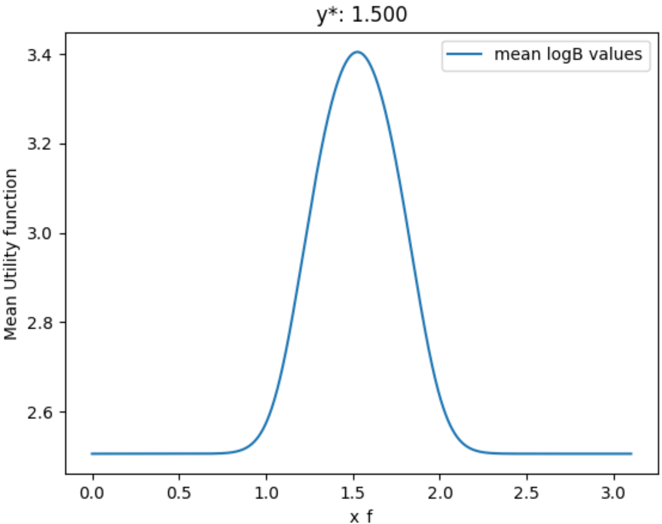
$y^* = 0$


y*: 0.000
mean logB values
Mean Utility function
2.675
2.650
2.625
2.600
2.575
2.550
2.525
2.500
0.0
0.5
1.0
1.5
2.0
2.5
3.0
x f

$y^* = 0.75$

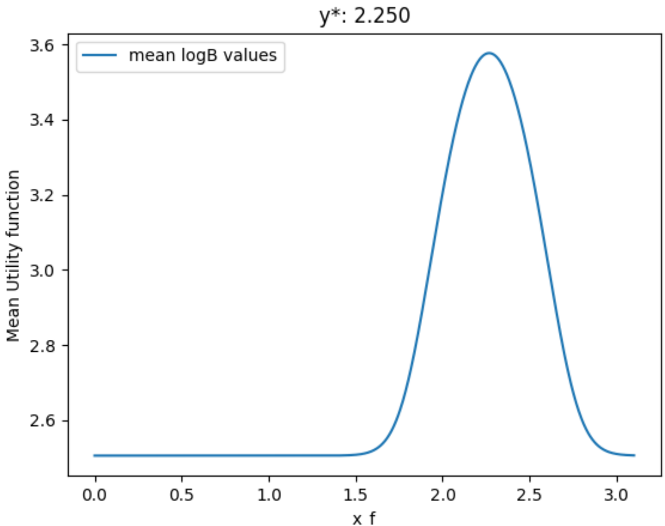

y*: 0.750
mean logB values
Mean Utility function
3.1
3.0
2.9
2.8
2.7
2.6
2.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
x f

$y^* = 1.5$



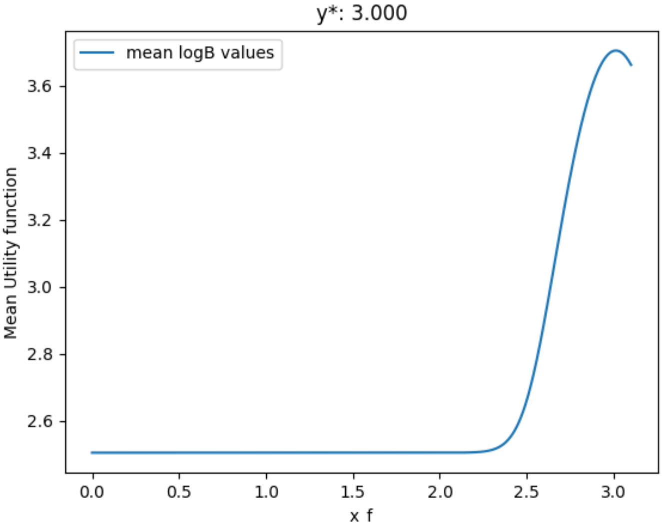
y*: 1.500
mean logB values
Mean Utility function
3.4
3.2
3.0
2.8
2.6
0.0
0.5
1.0
1.5
2.0
2.5
3.0
x f

$y^* = 2.25$



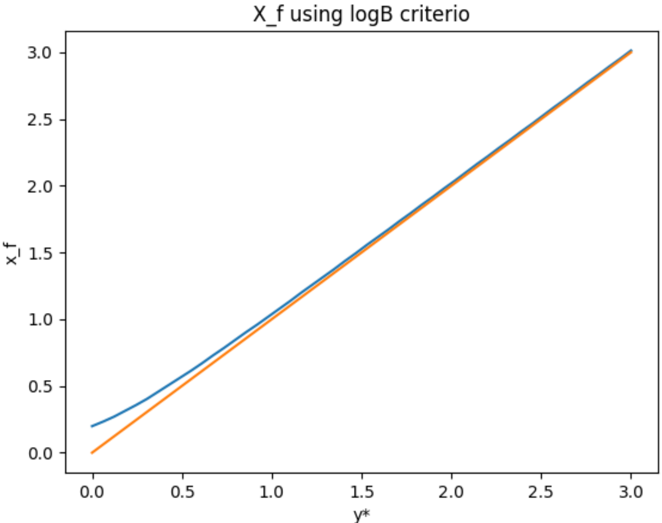
y*: 2.250
mean logB values
Mean Utility function
3.6
3.4
3.2
3.0
2.8
2.6
0.0
0.5
1.0
1.5
2.0
2.5
3.0
x f

$y^* = 3$

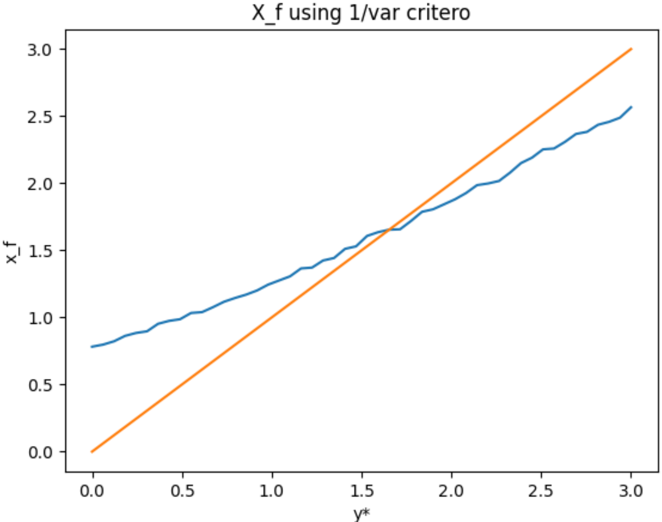


y*: 3.000
mean logB values
Mean Utility function
x f
3.6
3.4
3.2
3.0
2.8
2.6
0.0
0.5
1.0
1.5
2.0
2.5
3.0

x_f vrs y^* using the log Bayes factor



x_f vrs y^* using the inverse of the variance



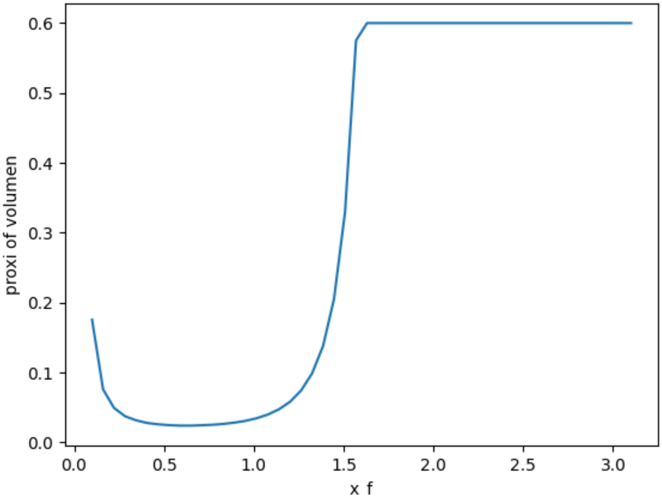
Working with the max volume of the hyper parameters

I estimate a proxi of the volume as the number of samples that have a value of the log Bayes greater than 2.5 with respect of a total number of samples, using Gibbs sampling

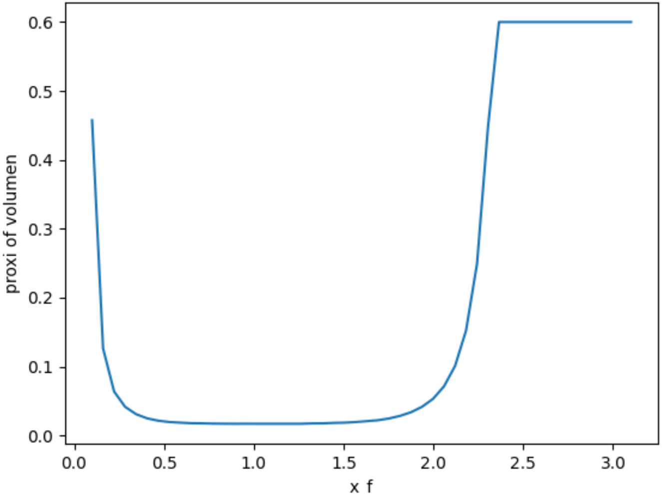
$$
\begin{aligned}
\log B &= \log p(d|M2) - \log p(d|M2) \\
&= \log \int_{\Sigma_2} p(d|M2,\theta_2)p(\theta|M2)d\theta_2 - \log \int_{\Sigma_1} p(d|M1,\theta_1)p(\theta|M1)d\theta_1 \\
&\approx \log \frac{1}{N}\sum_{i=1}^{N} p(d|M2,\theta_{2,i}) - \log \frac{1}{N}\sum_{i=1}^{N} p(d|M1,\theta_{1,i}), \quad \theta_{2,i} \sim p(\theta|M2), \quad \theta_{1,i} \sim p(\theta|M1)
\end{aligned}
\tag{16}
$$

Then, if I choose N $= 1$, would be equivalent to approximate logB integrated on a volume close to the samples. Then, I compute this for a number of times, and use the fraction of computations biguer than 2.5 as a proxi of the volume.

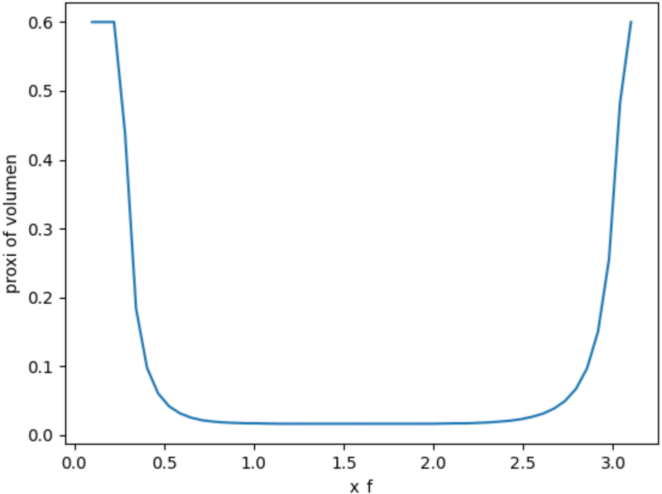
$y^* = 0$



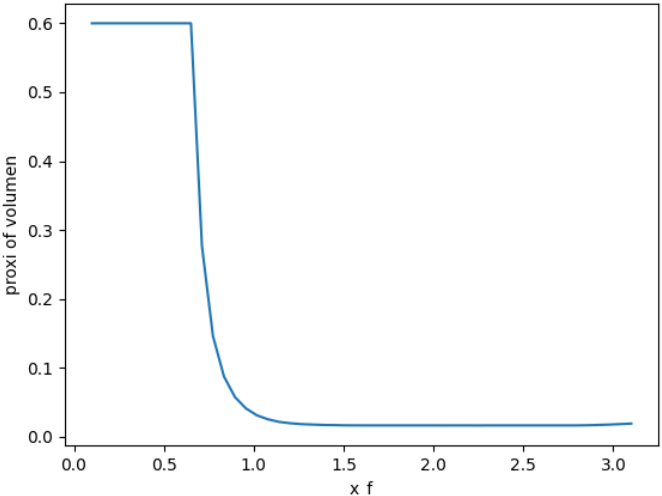
$y^* = 0.667$



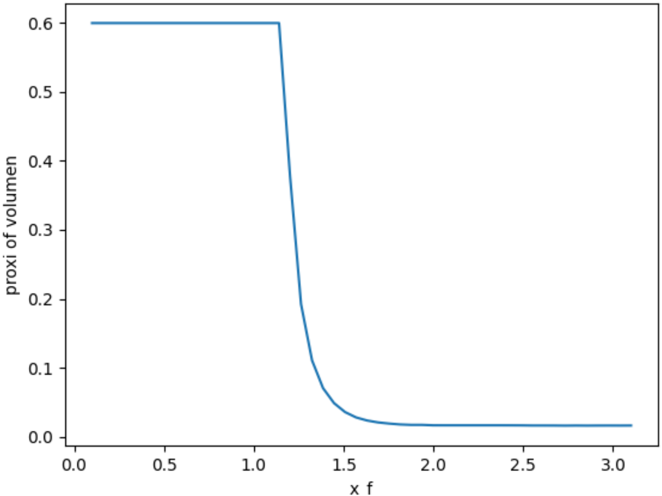
$y^* = 1.333$



$y^* = 2$



$y^* = 2.667$



The End