GRADUATE PROGRAMME IN STATISTICS AND OPERATIONAL RESEARCH
DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ATHENS

Bayesian Inference and Decision Theory
1st Project

Nikos Stamatis
nstam84@gmail.com

December 10, 2022

Contents

1 Analysis of the CG Content

1.1 Model Specification

We begin by formulating and studying the simplest Binomial model. Improvements and generalizations of this model are provided in the paragraphs that follow. Consider a DNA segment containing N bases at positions $j = 1, \ldots, N$ and let X_j denote the type of base in the j-th position. In particular, let $X_j = 1$ when the base is either C or G and $X_j = 0$ otherwise. In both of our models, we assume that the sequence $(X_j)_{j=1}^N$ is independent. Considering the distribution of each X_j, it is natural to assume a Bernoulli distribution with a parameter $p_j \in (0,1)$.

Suppose that we have n such consecutive, non overlapping segments and let $(X_j^i)_{j=1}^N$ be the sequence of base types that appear in the i-th segment. We define a sequence $(Y_i)_{i=1}^n$ as follows: $Y_i = \sum_{j=1}^N X_j^i$. Each Y_i counts the number of C or G appearances in the i-th segment and is the sum of N independent Bernoulli trials. In particular, the given data $(y_1, \ldots, y_{100})$ are a realization of such a sequence $(Y_i)_{i=1}^n$ for $N = 5000$ and $n = 100$.

Model M_1: We suppose that at every point, the probability of "drawing" a C or G base is the same. This means that for every $i = 1, \ldots, n$, the sequence $(X_j^i)_{j=1}^N$ is identically distributed with $X_j^i \sim \text{Bernoulli}(p)$. Consequently, the sequence $(Y_i)_{i=1}^n$ is also identically distributed with $Y_i \sim \text{Binomial}(N, p)$ and independent, since the Y_i's are finite sums on partitions of the same (independent) sequence $(X_j^i)_{i,j}$.

Model M_2: In this model we do not assume that all p_i's are the same, but rather that there are only two possible values for them, say p_1 and p_2. Moreover, we assume that there exists a $t' \in \{1, \ldots, N \cdot n\}$ such that all the X_i's before t follow a Bernoulli(p_1), whilst every X_i that appears after this position, follows a Bernoulli(p_2) distribution. In turn, this implies that there exists a $t \in \{1, \ldots, n\}$, such that $Y_1, \ldots, Y_t \sim \text{Binomial}(N, p_1)$, whilst $Y_{t+1}, \ldots, Y_N \sim \text{Binomial}(N, p_2)$.[†]

Before we proceed with the conjugate analysis of the two models, lets examine whether the diagrams (Figure 1) of our data support the existence of isochores that justify our assumptions: The first data set seems to come from a one isochore. The data do exhibit a large variation, but there is no indication of a drastic change in the mean value of the GC-content at any region. In the second diagram, we observe a significant change after $t = 40$, which indicates a second isochore starting from there on.

One can visually detect an increasing trend in the segment $[0, 50]$ of the third data set and a decreasing trend in the same segment of the fourth data set. Clearly, such trends suggest a certain degree of correlation between the random variables $(Y_i)_i$, which is incompatible with the independence assumption and they probably shouldn't be classified as isochore segments. In the final diagram there appear to be two different type isochores merging one within another, producing four consecutive isochores, two of each type, with endpoints near $t = 40, 60, 80$.

1.2 Conjugate Analysis

1.2.1 The One Isochore Model M_1

For every $i = 1, \ldots, n$, $Y_i \sim \text{Binomial}(N, p)$. We assume a Beta($a, b$) prior distribution for the parameter p. More specifically, having no prior information about the phenomenon in question, we choose the Jeffrey's prior Beta($\frac{1}{2}, \frac{1}{2}$). In what follows, we write the conjugate analysis of our model in its most general form and in the end, we substitute our particular choice of prior, as well as the constants $n = 100$ and $N = 5000$.

Let $\tilde{y} = (y_1, \ldots, y_n)$ be our sample. The density of one observation is $f(y_i \mid p) = \binom{N}{y_i} p^{y_i}(1-p)^{N-y_i}$, the joint density of the sample is

$$f(\tilde{y} \mid p) = \prod_{i=1}^{n} \binom{N}{y_i} \cdot p^{\sum_{i=1}^n y_i}(1-p)^{nN - \sum_{i=1}^n y_i},$$

[†]This is not entirely accurate: If the real change point t' occurs inside one of the subsegments of length N, say the k-th segment, then Y_k does not follow either one of the aforementioned distributions, but a mixture of them. We are willing to overlook this small inaccuracy.

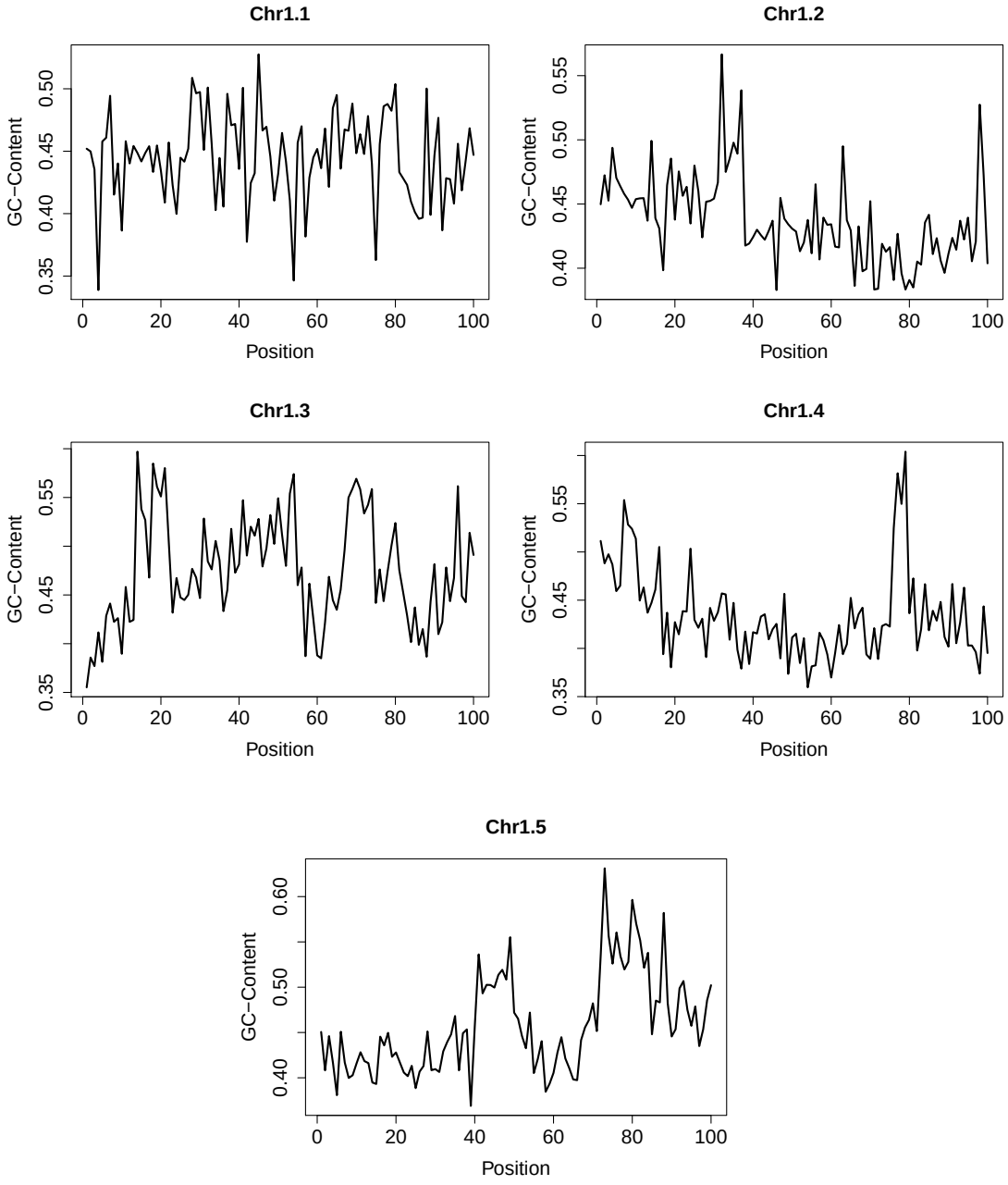

Figure 1: The GC-Content variability in each of the five data sets Chr1.1-Chr1.5.

whereas our prior distribution for p is $f(p) = \frac{1}{\mathrm{B}(a,b)} p^{a-1}(1-p)^{b-1}$. Therefore, the posterior density of p is $f(p \mid \tilde{y}) \propto p^{\sum_{i=1}^n y_i + a - 1}(1-p)^{nN - \sum_{i=1}^n y_i + b - 1}$, which is, up to a constant, equal to the density of $\mathrm{Beta}\left(a + \sum_{i=1}^n y_i, b + nN - \sum_{i=1}^n y_i\right)$. So the posterior distribution of p is

$$p \mid \tilde{y} \sim \mathrm{Beta}\left(a + \sum_{i=1}^n y_i, b + nN - \sum_{i=1}^n y_i\right). \tag{1.1}$$

The *evidence* of model M_1 is

$$\begin{aligned} f(\tilde{y} \mid M_1) &= \int_0^1 f(\tilde{y} \mid p) f(p)\, dp = \frac{\prod_{i=1}^n \binom{N}{y_i}}{\mathrm{B}(a,b)} \int_0^1 p^{\sum_{i=1}^n y_i} (1-p)^{nN-\sum_{i=1}^n y_i} p^{a-1}(1-p)^{b-1}\, dp \\ &= \frac{\prod_{i=1}^n \binom{N}{y_i}}{\mathrm{B}(a,b)} \int_0^1 p^{\sum_{i=1}^n y_i + a - 1} (1-p)^{nN+b-1-\sum_{i=1}^n y_i}\, dp \qquad (1.2) \\ &= \frac{\prod_{i=1}^n \binom{N}{y_i}}{\mathrm{B}(a,b)} \mathrm{B}\left(a + \sum_{i=1}^n y_i, b + nN - \sum_{i=1}^n y_i\right), \qquad (1.3) \end{aligned}$$

as the integrand in (1.2) is just the density of the $\mathrm{Beta}\left(a+\sum_{i=1}^n y_i, b+nN-\sum_{i=1}^n y_i\right)$ distribution, without the normalization constant.

1.2.2 The Two Isochore Model M_2

Let T denote the random variable describing the position of the change point, $T = 1,\ldots,n-1$ and $(q_t)_{t=1}^{n-1}$, be its distribution. We have no prior information on where the change point may have occurred, so we choose the Discrete Uniform distribution for T, that is $q_t = \frac{1}{n-1}$ for $t = 1,\ldots,n-1$. The observations $y_1,\ldots,y_t$ have a density of $f(y_i \mid p_1) = \binom{N}{y_i} p_1^{y_i}(1-p_1)^{N-y_i}$ and the observations $y_{t+1},\ldots,y_n$ have $f(y_i \mid p_2) = \binom{N}{y_i} p_2^{y_i}(1-p_2)^{N-y_i}$. The prior distributions for the parameters P_1 and P_2 are $\mathrm{Beta}(a_1,b_1)$ and $\mathrm{Beta}(a_2,b_2)$ respectively and we choose the Jeffrey's prior $\mathrm{Beta}(\frac{1}{2},\frac{1}{2})$ for both of them. Then

$$\begin{aligned} f(\tilde{y} \mid p_1,p_2,t) &= \prod_{i=1}^t \binom{N}{y_i} \cdot p_1^{\sum_{i=1}^t y_i}(1-p_1)^{tN-\sum_{i=1}^t y_i} \cdot \prod_{i=t+1}^n \binom{N}{y_i} \cdot p_2^{\sum_{i=t+1}^n y_i}(1-p_2)^{(n-t)N-\sum_{i=t+1}^n y_i} \\ &= \prod_{i=1}^n \binom{N}{y_i} \cdot p_1^{\sum_{i=1}^t y_i}(1-p_1)^{tN-\sum_{i=1}^t y_i} \cdot p_2^{\sum_{i=t+1}^n y_i}(1-p_2)^{(n-t)N-\sum_{i=t+1}^n y_i} \quad \text{and} \\ f(p_1,p_2,t \mid \tilde{y}) &\propto \prod_{i=1}^n \binom{N}{y_i} \cdot p_1^{\sum_{i=1}^t y_i}(1-p_1)^{tN-\sum_{i=1}^t y_i} \cdot p_2^{\sum_{i=t+1}^n y_i}(1-p_2)^{(n-t)N-\sum_{i=t+1}^n y_i} f(p_1) f(p_2) q_t \\ &\propto p_1^{\sum_{i=1}^t y_i + a_1 - 1}(1-p_1)^{tN-\sum_{i=1}^t y_i + b_1 - 1} \cdot p_2^{\sum_{i=t+1}^n y_i + a_2 - 1}(1-p_2)^{(n-t)N + b_2 - 1 - \sum_{i=t+1}^n y_i}. \end{aligned}$$

Assuming independence of T, P_1 and P_2, their joint density is the product of their marginal densities, $f(p_1,p_2,t) = f(p_1)f(p_2)q_t$, so the evidence of M_2 is

$$\begin{aligned} f(\tilde{y} \mid M_2) &= \int f(\tilde{y} \mid p_1,p_2,t) f(p_1,p_2,t)\, dp_1\, dp_2\, dt \\ &= \frac{\prod_{i=1}^n \binom{N}{y_i}}{\mathrm{B}(a_1,b_1)\,\mathrm{B}(a_2,b_2)} \cdot \sum_{t=1}^{n-1} q_t \int_0^1 \int_0^1 \Big[p_1^{\sum_{i=1}^t y_i}(1-p_1)^{tN-\sum_{i=1}^t y_i} p_2^{\sum_{i=t+1}^n y_i} \cdot \\ &\qquad \cdot (1-p_2)^{(n-t)N-\sum_{i=t+1}^n y_i} p_1^{a_1-1}(1-p_1)^{b_1-1} p_2^{a_2-1}(1-p_2)^{b_2-1} \Big] dp_1\, dp_2 \\ &= \frac{\prod_{i=1}^n \binom{N}{y_i}}{\mathrm{B}(a_1,b_1)\,\mathrm{B}(a_2,b_2)} \sum_{t=1}^{n-1} q_t\, \mathrm{B}\left(a_1 + \sum_{i=1}^t y_i, b_1 + Nt - \sum_{i=1}^t y_i\right) \mathrm{B}\left(a_2 + \sum_{i=t+1}^n y_i, b_2 + N(n-t) - \sum_{i=t+1}^n y_i\right). \end{aligned}$$

1.3 Model Comparison

We will compare the two models using the Bayes Factor BF_{12}. A value $\mathrm{BF}_{12} > 1$ indicates support for model M_1, whilst $\mathrm{BF}_{12} < 1$ indicates support for the second model. In terms of the logarithm of this quantity, the same conclusion can made according to whether $\ln \mathrm{BF}_{12}$ is negative or positive. Larger values of BF_{12} or $\ln \mathrm{BF}_{12}$, indicate stronger evidence in favor of the model M_1. To assign a more concrete meaning to these numbers, we follow the interpretation that is given in [KR95], as summarized in Table 1.

$$\begin{aligned}
\mathrm{BF}_{12} &= \frac{f(\tilde{y}\,|\,M_1)}{f(\tilde{y}\,|\,M_2)} \\
&= \frac{\mathrm{B}(a_1,b_1)\cdot\mathrm{B}(a_2,b_2)\cdot\mathrm{B}\left(a+\sum_{i=1}^{n} y_i,\ b+Nn-\sum_{i=1}^{n} y_i\right)}{B(a,b)\cdot\sum_{t=1}^{n-1} q_t\,\mathrm{B}\left(a_1+\sum_{i=1}^{t} y_i, b_2+Nt-\sum_{i=1}^{t} y_i\right)\mathrm{B}\left(a_2+\sum_{i=t+1}^{n} y_i,\ b_2+N(n-t)-\sum_{i=t+1}^{n} y_i\right)} \\
&= \frac{\pi(n-1)\,\mathrm{B}\left(0.5+\sum_{i=1}^{n} y_i,\ 0.5+Nn-\sum_{i=1}^{n} y_i\right)}{\sum_{t=1}^{n-1}\mathrm{B}\left(0.5+\sum_{i=1}^{t} y_i,\ 0.5+Nt-\sum_{i=1}^{t} y_i\right)\mathrm{B}\left(0.5+\sum_{i=t+1}^{n} y_i,\ 0.5+N(n-t)-\sum_{i=t+1}^{n} y_i\right)},
\end{aligned} \tag{1.4}$$

for the specific priors used in our model, as $\mathrm{B}(0.5,0.5)=\pi$ and $q_t = \frac{1}{n-1}$ for every $t = 1,\ldots,n-1$.

We computed (1.4) for the five data sets, by running the code in Listings 1 and 2 (pp. 12-13). Our results are presented in Table 2. As we can see, the evidence in favor of the Two Isochore model M_2 are, in all five cases, overpowering.

Table 1: An interpretation of the *Bayes Factor* BF_{10} of the hypothesis H_0 against H_1 [KR95, p. 776].

$2\ln \mathrm{BF}_{10}$	BF_{10}	Evidence against H_0
0-2	1-3	Barely worth mentioning
2-6	3-20	Positive
6-10	20-150	Strong
>10	>150	Very Strong

1.3.1 Can we detect a true* One Isochore?

In all five data sets, our method always chose the Two Isochore model. Before attempting to answer why this happened, we would like to make sure that our method can at least detect a blatantly true one isochore. We would also like to test its limitations and discover when it can be tricked into falsely detecting a changepoint. We performed two type of simulations to test how our method behaved when we applied it to regulated inputs.

In the first one (see Listing 3, p. 13), we drew ten thousand different samples following a Binomial distribution with the same mean as the first data set. All these samples belong to a true One Isochore Model and our method should be able to choose it over M_2. Indeed, only 0.06% of the time it falsely detected a changepoint and even when it did, the Bayes Factor was low enough not to suggest significant evidence in favor of it.

Table 2: In all five cases, the model M_2 was chosen and the exact change points of each one are mentioned in the third column. The last two columns contain the posterior probabilities at each of the two isochores.

Dataset	$\ln \mathrm{BF}_{12}$	Change point	p_1	p_2
Chr1.1	-46	80	0.448	0.430
Chr1.2	-417	37	0.465	0.422
Chr1.3	-621	13	0.410	0.484
Chr1.4	-516	16	0.487	0.425
Chr1.5	-1148	71	0.437	0.512

In the second one (see Listing 4, p. 13), ten thousand samples were drawn from a Beta-Binomial model with the same mean as the first

*And what exactly is a *true* isochore to begin with?

data set. The purpose of this simulation was to find out whether our method could be tricked[†] into falsely[†] identifying these draws as coming form a Two Isochore Model. This is exactly what happened in all of the ten thousand draws, however, the failure of our model in this instance should not come as a surprise. The method we developed has not been trained to detect true isochores, but rather isochores that come from the same Binomial distribution. If we change the distribution of Y_i's, then our method may misinterpret the deviance from a Binomial distribution as the existence of a second isochore.

In Paragraph 1.5, we provide a definition of what a "true" isochore is and we show that the collection of the distributions that can produce a true isochore is vast. Any model, like M_1 or M_2, that assumes a particular distribution from this class may fail to give a correct result if the data we provide it with, comes from a different distribution from the same class. In order to accurately detect isochores, ideally we would like to develop a method which would work for a diverse set of different distributions, or at the very least, select carefully the most appropriate distribution from this class.

1.4 The Issue with Overdispersion

In this section we address the issue with the overdispersion of our data. A look at Figure 1 reveals that the simple Binomial model we assumed, does not account for the variability of our data. This is more clearly depicted in Figure 3, p. 9, where we compare our five data sets to data drawn from the corresponding Binomial models. If the data sets had been drawn from a Binomial distribution, then they would exhibit a much larger variation than they currently do. Simulating from a Binomial Two Isochore Model, reveals the same issue (see Figure 2, p. 7). So, although the Two Isochore Model M_2 is much better than the One Isochore Model M_1, they are both quite poor in modelling the actual phenomenon.

1.4.1 Re-examining the Isochore Hypothesis

Isochores in models M_1 and M_2, had the additional property that the probability p of finding a Guanine or a Cytosine base in any particular *spot* was constant. This is an unnecessarily strong assumption and as we will see in Paragraph 1.5, there is a whole class of models which generalize this idea, by replacing the word "spot" with the word "region". Models M_1 and M_2 belong to this class, but they also happen to be its most simplistic representatives.

In this paragraph we examine how we can make a first step towards such a generalization: We adopt a more liberal definition of an isochore, as a region where the probability distribution P of finding a Guanine or a Cytosine base in any spot of a particular segment is the same. Under this definition, and by letting each P_i follow the same Beta distribution, the One and Two Isochore Models M_1, M_2 can be re-formulated as follows:

Model M_3: For every $i = 1,\dots,n$, $Y_i \,|\, P_i \sim \text{Binomial}(N, P_i)$ and $P_i \sim \text{Beta}(a,b)$ for some $a,b>0$.
Model M_4: There exists a $t \in \{1,\dots,n-1\}$ such that for every $i = 1,\dots n$, $Y_i \,|\, P_i \sim \text{Binomial}(N, P_i)$ and $P_i \sim \text{Beta}(a_1,b_1)$ for $i = 1,\dots,t$, whereas $P_i \sim \text{Beta}(a_2,b_2)$ for $i = t+1,\dots,n$, for some a_1, b_1, $a_2, b_2 > 0$.

This is the Beta-Binomial hierarchy and is commonly used when dealing with overdispersion of data that are suspected to belong to a Binomial distribution. As the following Lemma suggests, such a model does indeed increase the variance, while keeping the mean unaffected.

[†]The use of these words suggests that we view the Beta-Binomial model as a one isochore model. This is indeed the case (for a formal proof of this claim, see Corollary 1.5.4, p. 11).

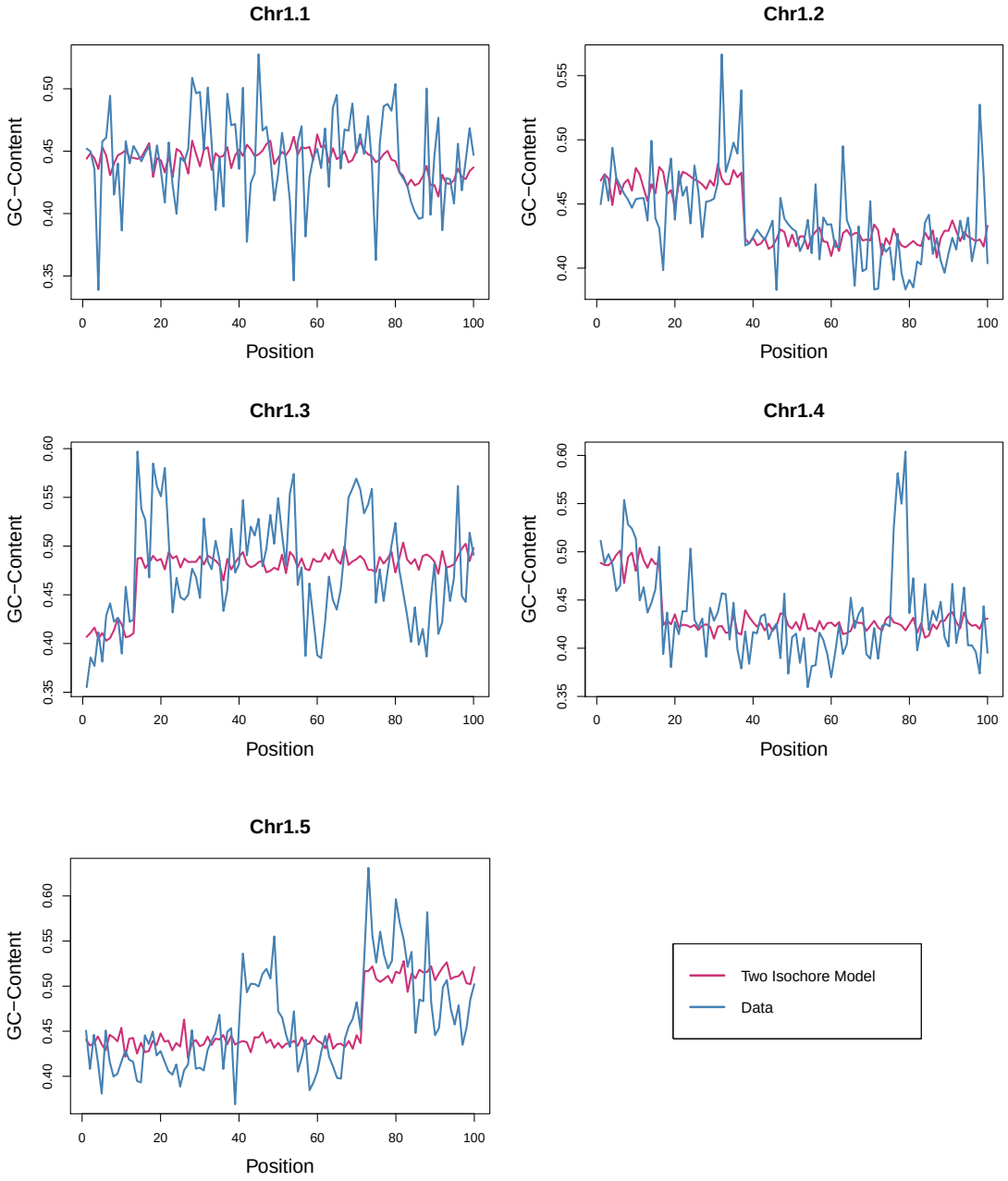

Figure 2: The GC-Content variability in each of the five data sets **Chr1.1-Chr1.5**, compared to data simulated from a true (Binomial) **Two Isochore Model**.

Lemma 1.4.1. *Let Z be a random variable with $Z \sim \text{Binomial}(N,p)$, $N \in \mathbb{N}$ and Y, P be random variables such that $Y \mid P \sim \text{Binomial}(N,P)$ and $P \sim \text{Beta}(a,b)$ with $p = \frac{a}{a+b}$. Then $\mathbb{E}[Y] = N\mathbb{E}[P] = Np$ and $\text{Var}\, Z = \frac{Nab}{(a+b)^2} < \text{Var}\, Y = \frac{Nab(a+b+N)}{(a+b)^2(a+b+1)}$.*

Proof. The proof is a simple exercise in probability (see [CB01, pp. 167-168] and is omitted. □

We proceed with the conjugate analysis of model M_3. Recall that according to M_3, $Y_i \mid P_i \sim$ Binomial(N, P_i) and $P_i \sim$ Beta(a,b). For $\tilde{y} = (y_1, \ldots, y_n)$ and $\tilde{p} = (p_1, \ldots, p_n)$,

$$f(\tilde{y} \mid \tilde{p}) = \prod_{i=1}^{n} \binom{N}{y_i} p_i^{y_i} (1-p_i)^{N-y_i}, \quad f(\tilde{p}) = \frac{1}{\mathrm{B}(a,b)^n} \prod_{i=1}^{n} p_i^{a-1}(1-p_i)^{b-1}, \text{ so}$$

$$\begin{aligned} f(\tilde{y} \mid M_3) &= \frac{\prod_{i=1}^{n} \binom{N}{y_i}}{\mathrm{B}(a,b)^n} \int \prod_{i=1}^{n} p_i^{y_i+a-1}(1-p_i)^{N-y_i+b-1} \, dp_1 \cdots dp_n \\ &= \frac{\prod_{i=1}^{n} \binom{N}{y_i}}{\mathrm{B}(a,b)^n} \prod_{i=1}^{n} \mathrm{B}(a+y_i, b+N-y_i). \end{aligned}$$

For the model M_4, $f(\tilde{y} \mid \tilde{p}, t)$ is the same as before, whilst

$$f(\tilde{p}, t) = \frac{q_t}{B(a_1,b_1)^t B(a_2,b_2)^{n-t}} \prod_{i=1}^{t} p_i^{a_1-1}(1-p_i)^{b_1-1} \prod_{i=t+1}^{n} p_i^{a_2-1}(1-p_i)^{b_2-1}, \text{ so}$$

$$\begin{aligned} f(\tilde{y} \mid M_4) = \prod_{i=1}^{n} \binom{N}{y_i} \sum_{t=1}^{n} \frac{q_t}{B(a_1,b_1)^t B(a_2,b_2)^{n-t}} \int \prod_{i=1}^{t} p_i^{y_i+a_1-1}(1-p_i)^{N-y_i+b_1-1} \cdot \\ \prod_{i=t+1}^{n} p_i^{y_i+a_2-1}(1-p_i)^{N-y_i+b_2-1} \, dp_1 \cdots dp_n \\ = \prod_{i=1}^{n} \binom{N}{y_i} \sum_{t=1}^{n} \frac{q_t}{B(a_1,b_1)^t B(a_2,b_2)^{n-t}} \prod_{i=1}^{t} \mathrm{B}(a_1+y_i, b_1+N-y_i) \prod_{i=t+1}^{n} B(a_2+y_i, b_2+N-y_i). \end{aligned}$$

Assuming the Jeffrey's priors, or any other prior where $a_1 = a_2$ and $b_1 = b_2$, the models M_3 and M_4 are equivalent. Since we have no prior information on either of these two distributions, we will refrain from comparing these two models to each other. However, we compared M_3 to M_2 using the code in Listing 5 at p. 14, and the results are shown in Table 3. We observe that M_3 is much better than M_2 in all five data sets.

1.5 The Class of Isochore Models

In the previous paragraph we may have given the wrong impression that the models M_3 and M_4 diverge from the philosophy of the exercise. One could argue that in an isochore, the probability should be constant and not drawn repeatedly from a distribution, therefore varying among different positions. However, this is definitely not the case: All four models examined so far, are equally faithful to the notion of the isochore, as introduced in the statement of the exercise.

Table 3: Comparing models M_2 and M_3.

Dataset	$\ln \mathrm{BF}_{23}$
Chr1.1	-775
Chr1.2	-324
Chr1.3	-2006
Chr1.4	-1199
Chr1.5	-1322

Recall that isochores are *regions where the proportion of C or G bases is roughly constant*. Both the Binomial and Beta-Binomial models lead to the existence of regions that satisfy exactly this property. In fact, any sequence of random variables $(Y_n)_n$ for which the Strong Law of Large Numbers holds, gives rise to an isochore and is a potential candidate to model our problem. We collect the random variables with this property into a class, as described below.

Definition 1.5.1. Let $(Y_n)_{n=1}^{\infty}$ be a a sequence of (not necessarily independent, nor identically distributed) random variables. We say that $(Y_n)_{n=1}^{\infty}$ has the *isochore property*, if there exists an $\mu > 0$ such that

$$\frac{Y_1 + \ldots + Y_n}{n} \longrightarrow \mu, \quad \text{almost surely.} \tag{1.5}$$

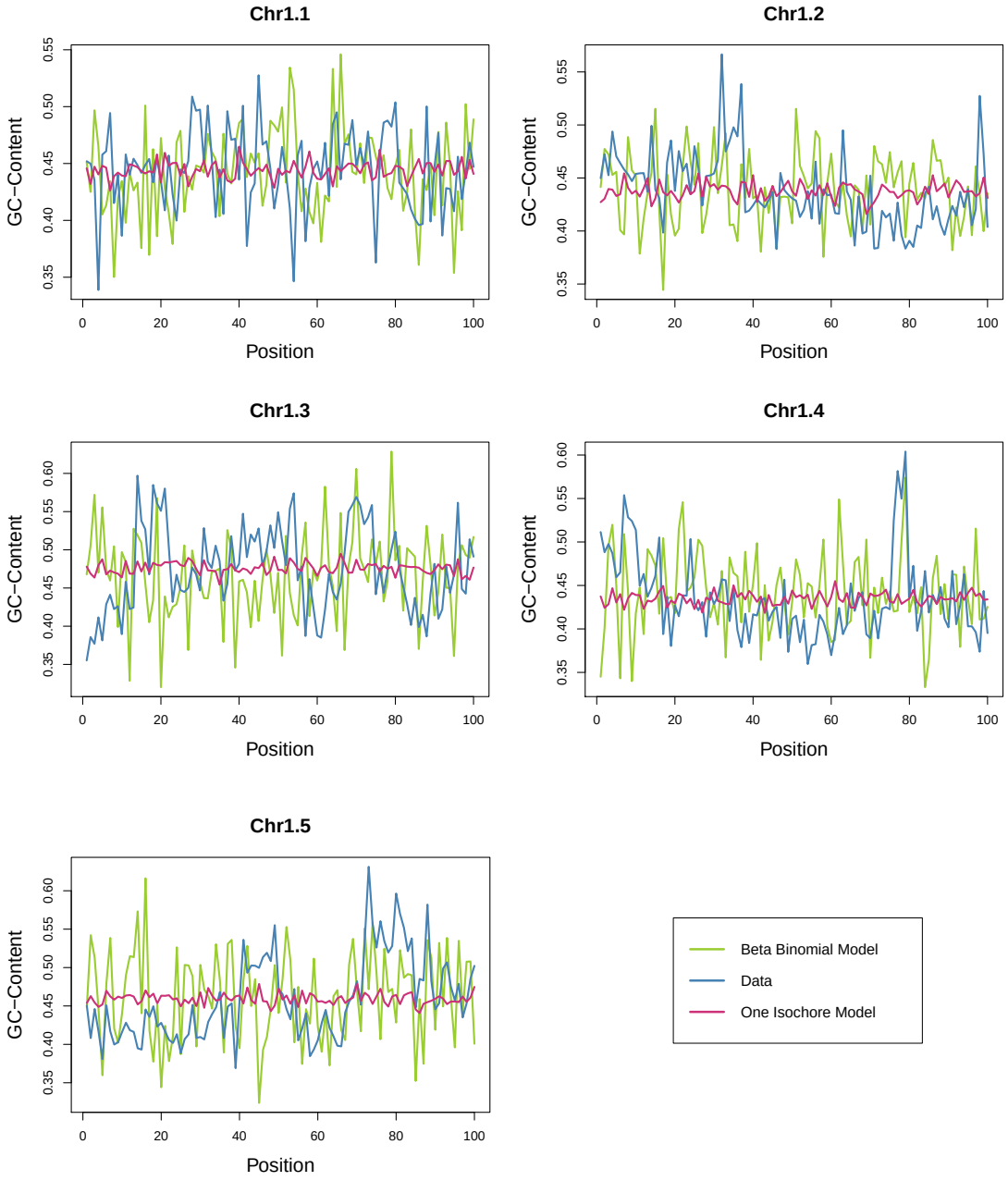

Figure 3: The GC-Content variability in each of the five data sets **Chr1.1-Chr1.5**, compared to data simulated from a true (Binomial) **One Isochore Model** and a **Beta-Binomial Model** with the same mean.

If each Y_n counts the number of occurrences of an event, then any sequence $(Y_n)_n$ that satisfies the definition above, also satisfies the definition of an isochore given in the statement of the exercise. The converse is also true. So the isochore notion, as described in the exercise, is just an intuitive interpretation of the conclusion of the SLLN. In particular, a sequence with the isochore property, generates a One Isochore model.

Given two such sequences, say $(Y_n)_n$ and $(Z_n)_n$, with respective limits $\mu_1 \neq \mu_2$, it is rather straightforward to merge them in a way to obtain a Two Isochore Model: Let $t, n \in \mathbb{N}$ such that $t < n$ and all three numbers $n, t, n-t$ are large enough for the SLLN approximations to hold

adequately. If we define $(W_i)_{i=1}^n$ as $W_i = Y_i$ for $i = 1,\dots,t$ and $W_i = Z_i$ for $i = t+1,\dots,n$, then $(W_i)_{i=1}^n$ generates a Two Isochore model.

As the assumptions in Definition 1.5.1 are quite general, the set of distributions that could potentially describe an isochore is huge. However, not all of them may be appropriate for any phenomenon. In particular, for the data sets of the exercise, distributions obtained from the Beta-Binomial hierarchy would presumably be more preferable, since they constitute a natural generalization of the Binomial model. We conclude this paragraph by showing that every sequence of random variables that follow a Beta-Binomial hierarchy, also satisfies the isochore property.

Regarding the Beta-Binomial models M_3, M_4, the sequence $(Y_n)_n$ was i.i.d. and the isochore property follows by a simple application of the Strong Law of Large Numbers. However, we would like to prove the isochore property as generally as possible, so we need to relax the identically distributed property. Our main tool will be another theorem by Kolmogorov:

Theorem (Kolmogorov). *Let $(Y_n)_{n=1}^\infty$ be an independent sequence of random variables with $\sigma_n^2 = \operatorname{Var} Y_n < \infty$ for every $n \in \mathbb{N}$ and let $S_n = Y_1 + \dots + Y_n$. Suppose that*

$$\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty \quad \textit{(Kolmogorov condition).} \tag{1.6}$$

Then $\frac{S_n - \mathbb{E}[S_n]}{n} \to 0$ almost surely.

Proof. See [Pet95, Theorem 6.7, p. 209]. □

Using Kolmogorov's Theorem, we can formulate a sufficient condition for a sequence $(Y_n)_n$ to have the isochore property:

Lemma 1.5.2. *Let $(Y_n)_{n=1}^\infty$ be a sequence of independent random variables with finite means $\mu_n = \mathbb{E}[Y_n]$ and variances $\sigma_n^2 = \operatorname{Var}[Y_n]$. Suppose that the sequence $(\mu_n)_{n=1}^\infty$ is Cesàro convergent to some $\mu < \infty$ and that $(\sigma_n^2)_n$ satisfies Kolmogorov's condition. Then $(Y_n)_{n=1}^\infty$ has the isochore property.*

Proof. By Kolmogorov's Theorem, $\frac{S_n - \mathbb{E}[S_n]}{n} \to 0$ almost surely. Additionally, since $(\mu_n)_{n=1}^\infty$ is Cesàro convergent, $\frac{\mathbb{E}[S_n]}{n} = \frac{\mu_1 + \dots + \mu_n}{n} \to \mu$ and the conclusion follows from the triangle inequality. □

Proposition 1.5.3. *Let $(Y_n)_n$ be an independent sequence for which there exists an $N \in \mathbb{N}$, such that $Y_n \le N$ for every n. If the sequence $(\mathbb{E}[Y_n])_n$ is Cesàro convergent, then $(Y_n)_n$ has the isochore property.*

Proof. Since the sequence $(Y_n)_n$ is bounded by N, $\operatorname{Var}[Y_n] \le \frac{N}{4}$, by the Popoviciu's Inequality,[†] so $\sum_{n=1}^\infty \frac{\sigma_n^2}{n^2} \le \frac{N}{4} \cdot \frac{\pi^2}{6} < \infty$ and the Kolmogorov condition is satisfied. By Kolmogorov's Theorem, we deduce that $(Y_n)_n$ satisfies the conclusion of the SLLN, so $(Y_n)_n$ has the isochore property. □

[†] The rather elementary Popoviciu's inequality, states that if a random variable $Y : \Omega \to [m, M]$ takes values on a bounded interval, then $\operatorname{Var} Y \le \frac{M-m}{4}$ (see [ABCD, p. 430]). Improvements of this upper bound also exist.

Corollary 1.5.4. *Let $(Y_n)_n$ be a sequence of independent random variables with $Y_n \mid P_n \sim$ Binomial(N, P_n) and $P_n \sim$ Beta(a_n, b_n). If the sequence $(\frac{a_n}{a_n+b_n})_n$ is Cesàro convergent, then $(Y_n)_n$ satisfies the isochore property. In particular, if $(\mathbb{E}[P_n])_n$ is convergent, or constant, then $(Y_n)_n$ satisfies the isochore property.*

Proof. As we saw in Lemma 1.4.1, $\mathbb{E}[Y_n] = N\frac{a_n}{a_n+b_n}$, which is Cesàro convergent by our assumptions. Additionally, the sequence $(Y_n)_n$ is bounded by N and the previous proposition is applicable. □

Remark 1.5.5. The conditions above can easily be modified in the case where N is not the same for all Y_n's, that is, if we consider the following hierarchy: $Y_n \mid P_n \sim$ Binomial(N_n, P_n), $P_n \sim$ Beta(a_n, b_n). The modified mean condition is that the sequence $(N_n\frac{a_n}{a_n+b_n})_n$ needs to be Cesàro convergent. However, the sequence $(\sigma_n^2)_n$ is no longer bounded, but rather $\sigma_n^2 \leq \frac{N_n}{4}$ for every $n \in \mathbb{N}$. So, a sufficient condition in order for the Kolmogorov condition to hold, is that $\sum_{n=1}^{\infty}\frac{N_n}{n^2} < \infty$.

1.6 Conclusions

We analyzed the data under multiple different distribution assumptions, but always respecting the isochore notion. At first, we introduced two simple Binomial models M_1 (one isochore) and M_2 (two isochore). In all five data sets the evidence in favor of the two isochore model M_2 were overwhelming. The exact positions of the suggested changepoints were also provided.

However, none of these two models accurately modelled the phenomenon in question. We showed that the isochores in our data do not follow a Binomial distribution, as they exhibit overdispersion. Additionally, we explained the reasons why a Binomial model is not realistic, even at the early stages of building it, without having looked at the data. A Beta-Binomial model M_3 was suggested as a possible answer to both of these issues. Our analysis showed that M_3 was overwhelmingly better than the previous two.

Finally, we approached the question a little more abstractly. We showed that there exists a huge class of random variables, capable of generating one and two isochores in a very natural way. Restricted to the DNA example of the exercise, the subclass consisting of random variables built according to a Beta-Binomial hierarchy seem more preferable. They provide a more realistic approach to the problem and they are a natural extension of the simplistic Binomial models M_1 and M_2. They also capture the behavior of the data sets much better. But, as shown in Corollary 1.5.4, even this subclass is still vast, so a careful model specification is needed here as well.

1.7 Code used in R

Listing 1: The function `bfactor` in R, used to compute $\ln \mathrm{BF}_{12}$, the positions of the change point and the probabilities in each of the two isochores for any given dataset and parameters for the priors.

```
1  # bfactor computes lnBF, the position of the changepoint and
2  # the probabilities in each of the (possible) two isochores,
3  # given the data set and the prior distributions
4
5  bfactor<-function(data, a=0.5, b=0.5, a1=0.5,
6      b1=0.5, a2=0.5, b2=0.5, N=5000) {
7  mat<-numeric(4)
8  n=length(data); n1<-n-1; s<-rep(0, n1);
9  for (i in 1:n1) s[i]=sum(data)
10 schr=sum(data)
11 s<-cumsum(data)   # Partial sums of the data
12 s<-s[c(1:n1)]     # excluding the last one
13 schr=sum(data)
14 p1=0; p2=0; # Probabilities in each of the isochore segments
15 logpr<-rep(0,n1)
16 for (i in 1:n1){logpr[i]=lbeta(a1+s[i],   # Logarithms of
17     b1+N*i-s[i])+lbeta(a2+schr-s[i], # the quantities
18     b2+N*(n-i)-schr+s[i])}   # that appear in the evidence
19 logprmax=max(logpr)           # of M_2
20
21 LogEv2=-lbeta(a1,b1)-lbeta(a2,b2)-log(n-1)+logprmax+
22     log(sum(exp(logpr-logprmax)))
23 LogEv1=-lbeta(a,b)+lbeta(a+schr, b+N*n-schr)
24 LogEv1-LogEv2          # The logarithm of the Bayes Factor BF_12
25 pr=exp(logpr-logprmax)
26 pr=pr/sum(pr)
27 t=1:n1          # Position in the region of the chromosome
28 chp=t[pr==max(pr)][1] # Position of the possible changepoint
29 max(pr)
30 p1= (a1+sum(data[c(1:chp)]))/(a1+b1+N*chp) # Post. pr. in each
31 p2= (a2+sum(data[c((chp+1) : n)]))/(a2+b2+N*(n-chp)) # isochore
32 mat<-c(LogEv1-LogEv2, chp,p1,p2)
33 return(mat) }
```

Line 21: The logarithm of the Bayes Factor has the form $\ln \mathrm{BF}_{12} = \ln \frac{a}{\sum_{t=1}^{n-1} b_t}$, where a and b_t are given by (1.4). Due to computational limitations, it is impossible to estimate $\ln \sum_{t=1}^{n-1} b_t$ in R directly, so we use the following trick: Let $b_k = \max_{t=1,\ldots,n-1}\{b_t\}$. Then

$$\ln \sum_{t=1}^{n-1} b_t = \ln \sum_{t=1}^{n-1} b_t - \ln b_k + \ln b_k = \ln \sum_{t=1}^{n-1} \frac{b_t}{b_k} + \ln b_k = \ln \sum_{t=1}^{n-1} e^{\ln b_t - \ln b_k} + \ln b_k$$

and we compute the RHS of this expression instead.

Line 28: If the maximum of the vector `pr` is attained at more than one points, then the positions of these maxima `t[pr==max(pr)]` form a vector and, by adding `[1]` we choose the first element of it. This was not necessary when analyzing the data sets of the exercise, since the maxima were attained at a unique point in all five cases. However, due to the large number of the simulations, there were always generated data where this vector contained more than one element, causing obvious problems to the numerical manipulations that followed.

Listing 2: Code and results of running `bfactor` on each of the five datasets of the exercise.

```
analysis<-matrix (ncol=4, nrow=5)
analysis[1,]<-bfactor(chr1.1)
analysis[2,]<-bfactor(chr1.2); analysis[3,]<-bfactor(chr1.3)
analysis[4,]<-bfactor(chr1.4); analysis[5,]<-bfactor(chr1.5)
analysis
       ln BF  t    p1    p2
Chr1.1   -46 80 0.448 0.430
Chr1.2  -417 37 0.465 0.422
Chr1.3  -621 13 0.410 0.484
Chr1.4  -516 16 0.487 0.425
Chr1.5 -1148 71 0.437 0.512
```

Listing 3: A simulation consisting of applying our method to 10.000 draws from a Binomial sample resembling the first data set. Only 0.06% of the times it led to falsely detecting a changepoint.

```
simulateBeta11<-function(k) {
    SimResults=matrix(ncol=4, nrow=k)
    for (i in 1: k) {
            draw<- rbinom(100, 5000, mean(chr1.1/5000))
            SimResults[i,]<-bfactor(data=draw)
            }
    return (SimResults) }

res<-simulateBeta11(10000)
summary(res[,1])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -2.092   4.760   5.163   4.954   5.383   5.670
sum(res[,1]<0)/10000
[1] 6e-04
```

Listing 4: A simulation consisting of applying our method to 10.000 draws from a Beta-Binomial sample resembling the first data set, always led to rejecting the one isochore hypothesis.

```
BetaParam <- function(mu, var) { # Obtaining the Beta parame-
    param<-numeric(2)            # ters a, b, given the mean
    alpha <-(1-mu)*(mu^2)/var-mu  # mu and variance var
    bbeta <- mu*(1-mu)^2/var-1+mu
    param[1]<- alpha; param[2]<- bbeta
    return(param) }

simulateBetaBinom11<-function(k){
    SimResults=matrix(ncol=4, nrow=k)
    parameters<-BetaParam(mean(chr1.1/5000), var(chr1.1/5000))
    probab<-rbeta(100,  parameters[1], parameters[2])
    for (i in 1: k) {
        draw<-  rbinom(100, 5000,probab)
        SimResults[i,]<-bfactor(data=draw) }
    return (SimResults) }

resBB<-simulateBetaBinom11(10000)
summary(resBB[,1])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-106.50  -63.28  -56.06  -56.53  -49.13  -21.32
```

Listing 5: The function `bbfactor` used to compute $\ln \mathrm{BF}_{23}$. It also computes the evidence of model M_4, a Two Isochore Beta-Binomial model.

```
# bbfactor computes lnBF when comparing a Two Isochore
# Binomial model with a Beta Binomial model

bbfactor<-function(data, a=0.5, b=0.5, a1=0.5,
            b1=0.5, a2=0.5, b2=0.5, N=5000) {
mat<-numeric(5)
n=length(data); n1<-n-1; s<-rep(0, n1); betamat<-numeric(n)
s<-cumsum(data)
s<-s[c(1:n1)]
schr=sum(data)
p1=0; p2=0;
logpr<-rep(0,n1)
for (i in 1:n1){logpr[i]=lbeta(a1+s[i], b1+N*i-s[i])+
           lbeta(a2+schr-s[i], b2+N*(n-i)-schr+s[i])}
logprmax=max(logpr)

LogEv2=-lbeta(a1,b1)-lbeta(a2,b2)-log(n-1)+logprmax+
    log(sum(exp(logpr-logprmax)))
LogEv1=-lbeta(a,b)+lbeta(a+schr, b+N*n-schr)

for (i in 1:length(betamat)) {
    betamat[i]=lbeta(data[i]+a, N-data[i]+b)}

LogEv3=-n*lbeta(a, b)+ sum(betamat)

aux1<-numeric(n); aux2<-numeric(n); logppp<-numeric(n1)
for (i in 1:n) aux1[i]=lbeta(a1+data[i], N+b1-data[i])
        -lbeta(a1,b1)
for (i in 1:n) aux2[i]=lbeta(a2+data[i], N+b2-data[i])
        -lbeta(a2,b2)
for (i in 1:n1) logppp[i]=sum(aux1[c(1:i)])+sum(aux2[c((i+1):n)])
logpppmax=max(logppp)

LogEv4= -log(n-1)+log(sum(exp(logppp-logpppmax)))+logpppmax
mat<-c(LogEv1, LogEv2, LogEv3, LogEv4, LogEv2-LogEv3)
return(mat)}
```

2 A Simple Regression Model

Before starting with our analysis, we plotted the data of the exercise in Figure 4. We can clearly detect a linear relation between X and Y. This can also be established with the usual (frequentist) approach of the linear model $y = \alpha + \beta x + \varepsilon$. We ran a basic analysis in R, which concluded that both α and β are significant ($p < 2.2 \times 10^{-16}$) and estimated as $\hat{\alpha} = 2.05$ and $\hat{\beta} = 0.8$. The coefficient of determination of the model was $R^2 = 0.96$.

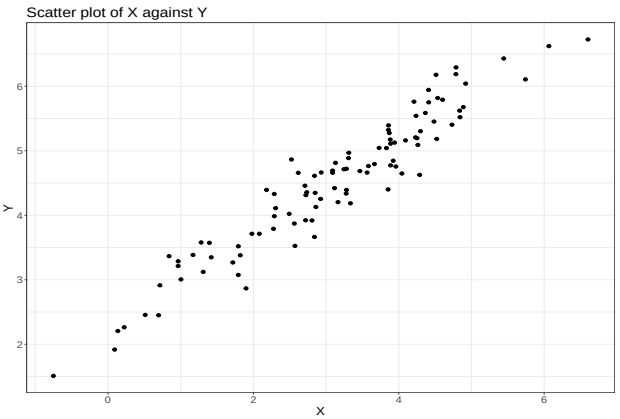

Figure 4: Scatter plot of the dataset `RegressionData.R`.

After examining the classical approach, let us focus on the the bayesian model. The joint prior distribution for α, β and τ is $f(\alpha,\beta,\tau) = f(\alpha,\beta\,|\,\tau)f(\tau)$, where both $f(\tau)$ and $f(\alpha,\beta\,|\,\tau)$ are known. For given α, β, τ and x, $y \sim$Normal$(\alpha+\beta x, \tau^{-1})$, so the distribution of the sample is given by

$$f(\tilde{y}\,|\,\alpha,\beta,\tau) = \frac{\tau^{\frac{n}{2}}}{\sqrt{2\pi}^n} e^{-\frac{\tau}{2}\sum_{i=1}^n (y_i - a - \beta x_i)^2}. \tag{2.1}$$

Since our goal is to examine whether $\beta = 0$ or $\beta \neq 0$, we will define two different models accordingly.

2.1 Conjugate Analysis

We begin with the analysis of the simplest model M_0, where $\beta = 0$. In this case

$$\begin{aligned} f(\tilde{y}\,|\,\alpha,\tau)f(\alpha,\tau) &= \frac{\tau^{\frac{n}{2}}}{\sqrt{2\pi}^n} e^{-\frac{\tau}{2}\sum_{i=1}^n (y_i-\alpha)^2} \frac{\sqrt{\tau c_1}}{\sqrt{2\pi}} e^{-\frac{c_1\tau\alpha^2}{2}} \frac{q^p}{\Gamma(p)} \tau^{p-1} e^{-q\tau} \\ &= \frac{\sqrt{c_1}q^p}{\sqrt{2\pi}^{n+1}\Gamma(p)} \tau^{p+\frac{n}{2}-\frac{1}{2}} \cdot e^{-\tau\left(\frac{1}{2}\sum_{i=1}^n (y_i-\alpha)^2 + q + \frac{c_1\alpha^2}{2}\right)} \end{aligned}$$

and the evidence of M_0 is

$$f(\tilde{y}\,|\,M_0) = \frac{\sqrt{c_1}q^p}{\sqrt{2\pi}^{n+1}\Gamma(p)} \int_0^\infty \underbrace{\int_{-\infty}^\infty \tau^{p+\frac{n}{2}-\frac{1}{2}} \cdot e^{-\tau\left(\frac{1}{2}\sum_{i=1}^n (y_i-\alpha)^2 + q + \frac{c_1\alpha^2}{2}\right)} d\alpha}_{I(\tau)}\, d\tau, \tag{2.2}$$

$$\begin{aligned} I(\tau) &= \tau^{p+\frac{n}{2}-\frac{1}{2}} \cdot \int_{-\infty}^\infty e^{-\tau\left(\frac{1}{2}\sum_{i=1}^n (y_i-\alpha)^2 + q + \frac{c_1\alpha^2}{2}\right)} d\alpha \\ &= \tau^{p+\frac{n}{2}-\frac{1}{2}} e^{-\tau(q+\frac{1}{2}\sum_{i=1}^n y_i^2)} \cdot \int_{-\infty}^\infty e^{-\frac{\alpha^2}{2}(n\tau+c_1\tau)} e^{\alpha\tau\sum_{i=1}^n y_i} d\alpha. \end{aligned}$$

The last integrand is just the density of a Normal distribution with mean $\mu = \frac{\sum_{i=1}^n y_i}{n+c_1}$ and variance $\sigma^2 = \frac{1}{\tau(n+c_1)}$, missing the term $\sigma\sqrt{2\pi}\cdot e^{\frac{\mu^2}{2\sigma^2}}$, so

$$\begin{aligned} I(\tau) &= \tau^{p+\frac{n}{2}-\frac{1}{2}} e^{-\tau\left(q+\frac{1}{2}\sum_{i=1}^n y_i^2\right)} \frac{\sqrt{2\pi}}{\sqrt{\tau(n+c_1)}} e^{\frac{\tau\left(\sum_{i=1}^n y_i\right)^2}{2(n+c_1)}} \\ &= \frac{\sqrt{2\pi}}{\sqrt{n+c_1}} \tau^{p+\frac{n}{2}-1} e^{-\tau\left(q-\frac{\left(\sum_{i=1}^n y_i\right)^2}{2(n+c_1)}+\frac{1}{2}\sum_{i=1}^n y_i^2\right)}. \end{aligned}$$

By integrating $I(\tau)$ with respect to τ, we have that

$$\begin{aligned} \int_0^\infty I(\tau)d\tau &= \frac{\sqrt{2\pi}}{\sqrt{n+c_1}} \int_0^\infty \tau^{p+\frac{n}{2}-1} e^{-\tau\left(q-\frac{\left(\sum_{i=1}^n y_i\right)^2}{2(n+c_1)}+\frac{1}{2}\sum_{i=1}^n y_i^2\right)} d\tau \\ &= \frac{\sqrt{2\pi}}{\sqrt{n+c_1}} \cdot \frac{\Gamma(p+\frac{n}{2})}{\left(q-\frac{\left(\sum_{i=1}^n y_i\right)^2}{2(n+c_1)}+\frac{1}{2}\sum_{i=1}^n y_i^2\right)^{p+\frac{n}{2}}}. \end{aligned} \tag{2.3}$$

We substitute (2.3), into (2.2) to obtain

$$f(\tilde{y}\,|\,M_0) = \frac{\sqrt{c_1}q^p}{\sqrt{n+c_1}\sqrt{2\pi}^n\Gamma(p)} \cdot \frac{\Gamma(p+\frac{n}{2})}{\underbrace{\left(q-\frac{\left(\sum_{i=1}^n y_i\right)^2}{2(n+c_1)}+\frac{1}{2}\sum_{i=1}^n y_i^2\right)^{p+\frac{n}{2}}}_{C}}. \tag{2.4}$$

Computing the evidence of the full model M_1 requires additional work, as we also need to integrate out β. As we saw previously,

$$\begin{aligned} f(\tilde{y}\,|\,\alpha,\beta,\tau)f(\alpha,\beta,\tau) &= \frac{\tau^{\frac{n}{2}}}{\sqrt{2\pi}^n} e^{-\frac{\tau}{2}\sum_{i=1}^n (y_i-\alpha-\beta x_i)^2} \frac{\sqrt{\tau c_1}}{\sqrt{2\pi}} e^{-\frac{c_1\tau\alpha^2}{2}} \frac{\sqrt{\tau c_2}}{\sqrt{2\pi}} e^{-\frac{c_2\tau\beta^2}{2}} \frac{q^p}{\Gamma(p)} \tau^{p-1} e^{-q\tau} \\ &= \frac{\sqrt{c_1c_2}q^p}{\sqrt{2\pi}^{n+2}\Gamma(p)} \tau^{p+\frac{n}{2}} \cdot e^{-\tau\left(\frac{1}{2}\sum_{i=1}^n (y_i-\alpha-\beta x_i)^2+q+\frac{c_1\alpha^2}{2}+\frac{c_2\beta^2}{2}\right)}. \end{aligned}$$

The evidence of M_1 is

$$\begin{aligned} f(\tilde{y}\,|\,M_1) &= \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty f(\tilde{y}\,|\,\alpha,\beta,\tau)f(\alpha,\beta,\tau)d\alpha d\beta d\tau \\ &= \frac{\sqrt{c_1c_2}q^p}{\sqrt{2\pi}^{n+2}\Gamma(p)} \int_0^\infty \int_{-\infty}^\infty \underbrace{\int_{-\infty}^\infty \tau^{p+\frac{n}{2}} \cdot e^{-\tau\left(\frac{1}{2}\sum_{i=1}^n (y_i-\alpha-\beta x_i)^2+q+\frac{c_1\alpha^2}{2}+\frac{c_2\beta^2}{2}\right)} d\alpha}_{I(\beta,\tau)} d\beta d\tau, \end{aligned}$$

$$\begin{aligned} I(\beta,\tau) &= \tau^{p+\frac{n}{2}} e^{-\tau(q+\frac{c_2\beta^2}{2}+\frac{1}{2}\sum_{i=1}^n (y_i-\beta x_i)^2)} \int_{-\infty}^\infty e^{-\alpha^2\frac{\tau(c_1+n)}{2}} e^{\alpha\tau\sum_{i=1}^n (y_i-\beta x_i)} d\alpha \\ &= \tau^{p+\frac{n}{2}} e^{-\tau(q+\frac{c_2\beta^2}{2}+\frac{1}{2}\sum_{i=1}^n (y_i-\beta x_i)^2)} \frac{\sqrt{2\pi}}{\sqrt{\tau(c_1+n)}} e^{\frac{\tau\left(\sum_{i=1}^n y_i-\beta x_i\right)^2}{2(n+c_1)}} \\ &= \frac{\sqrt{2\pi}}{\sqrt{\tau(c_1+n)}} \tau^{p+\frac{n}{2}-\frac{1}{2}} e^{-\tau\left(q+\frac{c_2\beta^2}{2}+\frac{1}{2}\sum_{i=1}^n (y_i-\beta x_i)^2-\frac{\left(\sum_{i=1}^n y_i-\beta x_i\right)^2}{2(n+c_1)}\right)}. \end{aligned} \tag{2.5}$$

By integrating $I(\beta,\tau)$ with respect to β, we have that

$$\int_{-\infty}^{\infty} I(\beta,\tau)d\beta = \frac{\sqrt{2\pi}}{\sqrt{\tau(c_1+n)}}\tau^{p+\frac{n}{2}-\frac{1}{2}}e^{-\tau\left(q+\frac{1}{2}\sum_{i=1}^n y_i^2-\frac{\left(\sum_{i=1}^n y_i\right)^2}{2(n+c_1)}\right)}\cdot$$

$$\cdot\int_{-\infty}^{\infty} e^{-\frac{\beta^2}{2}\left(\tau c_2+\tau\sum_{i=1}^n x_i^2-\frac{\tau\left(\sum_{i=1}^n x_i\right)^2}{n+c_1}\right)}e^{\beta\left(\tau\sum_{i=1}^n x_i y_i-\frac{\tau\sum_{i=1}^n x_i\sum_{i=1}^n y_i}{n+c_1}\right)}d\beta$$

$$= \frac{2\pi}{\sqrt{c_1+n}}\left(c_2+\sum_{i=1}^n x_i^2-\frac{\left(\sum_{i=1}^n x_i\right)^2}{n+c_1}\right)^{-\frac{1}{2}}\cdot$$

$$\cdot\underbrace{\tau^{p+\frac{n}{2}-\frac{1}{2}}\exp\left\{-\tau\left(q+\frac{1}{2}\sum_{i=1}^n y_i^2-\frac{\left(\sum_{i=1}^n y_i\right)^2}{2(n+c_1)}-\frac{1}{2}\frac{\left(\sum_{i=1}^n x_i y_i-\frac{\sum_{i=1}^n x_i\sum_{i=1}^n y_i}{n+c_1}\right)^2}{c_2+\sum_{i=1}^n x_i^2-\frac{\left(\sum_{i=1}^n x_i\right)^2}{n+c_1}}\right)\right\}}_{I(\tau)}.$$

The integral of $I(\tau)$ is

$$\int_0^{\infty} I(\tau)d\tau = \frac{\Gamma\left(p+\frac{n}{2}\right)}{\left(q+\frac{1}{2}\sum_{i=1}^n y_i^2-\frac{\left(\sum_{i=1}^n y_i\right)^2}{2(n+c_1)}-\frac{1}{2}\frac{\left(\sum_{i=1}^n x_i y_i-\frac{\sum_{i=1}^n x_i\sum_{i=1}^n y_i}{n+c_1}\right)^2}{c_2+\sum_{i=1}^n x_i^2-\frac{\left(\sum_{i=1}^n x_i\right)^2}{n+c_1}}\right)^{p+\frac{n}{2}}}, \tag{2.6}$$

so the evidence of M_1 is

$$f(\tilde{y}\,|\,M_2) = \frac{\sqrt{c_1c_2}q^p}{\sqrt{n+c_1}\sqrt{2\pi}^n\Gamma(p)}\underbrace{\left(c_2+\sum_{i=1}^n x_i^2-\frac{\left(\sum_{i=1}^n x_i\right)^2}{n+c_1}\right)^{-\frac{1}{2}}}_{A}\cdot$$

$$\cdot\underbrace{\frac{\Gamma\left(p+\frac{n}{2}\right)}{\left(q+\frac{1}{2}\sum_{i=1}^n y_i^2-\frac{\left(\sum_{i=1}^n y_i\right)^2}{2(n+c_1)}-\frac{1}{2}\frac{\left(\sum_{i=1}^n x_i y_i-\frac{\sum_{i=1}^n x_i\sum_{i=1}^n y_i}{n+c_1}\right)^2}{c_2+\sum_{i=1}^n x_i^2-\frac{\left(\sum_{i=1}^n x_i\right)^2}{n+c_1}}\right)^{p+\frac{n}{2}}}}_{B}. \tag{2.7}$$

Under the notation introduced in (2.4) and (2.7), the Bayes Factor BF_{01} is

$$\mathrm{BF}_{01} = \frac{f(\tilde{y}\,|\,M_0)}{f(\tilde{y}\,|\,M_1)} = \frac{AB}{\sqrt{c_2}C}. \tag{2.8}$$

2.2 Model Comparison

To compare the models M_0 and M_1, we computed the Bayes Factor BF_{01} using the code in Listing 6 and the result was $\mathrm{BF}_{01} = 6\cdot10^{-63}$. We can compute the posterior probabilities of the models M_0 and M_1, but the Bayes Factor is so small that the evidence in favor of the full model are overwhelming.

To further test the performance of our code, we executed 10.000 simulations. In each simulation, a value of τ was drawn from a Gamma(0.1, 0.1) distribution, a value of α from a

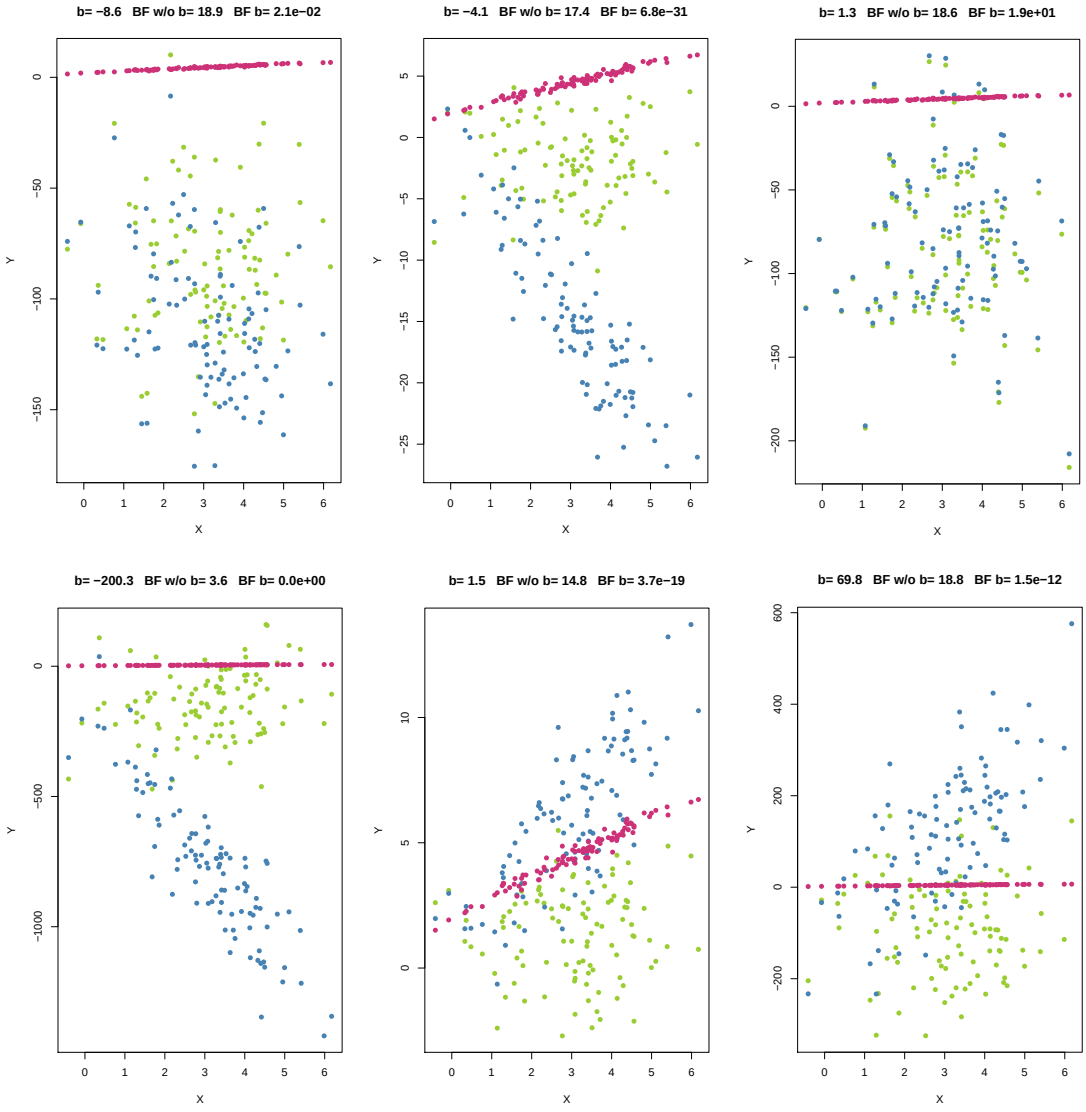

Figure 5: Six different simulations based on the model $y_i = \alpha + \beta x_i + \varepsilon_i$. The **green data** come from simulations where $\beta = 0$, whereas the **blue data** come from the full model. We included the scatter plot of the actual data set `RegressionData`, as well as the values of β and the Bayes Factor BF_{01}, when the data come from the green or the blue data set respectively.

Normal$(0,(0.5\tau)^{-1})$, a value of β from the same distribution and a vector $\tilde{\varepsilon} = (\varepsilon_1,\ldots,\varepsilon_{100})$, where each ε_i was drawn from a Normal$(0,\tau^{-1})$. Then we constructed two data sets $(y_n)_{n=1}^{100}$ and $(z_n)_{n=1}^{100}$, where each y_n and z_n was defined as $y_n = a + \varepsilon_n$ and $y_n = a + \beta x_n + \varepsilon_n$. Finally, we computed the Bayes Factor BF_{01} for each of these two data sets. Assuming the correctness of our model, we would expect it to generally give a high value of BF_{01} for the first data set and a low for the second one.

The results indicate that our method works well. When the data were coming from the model for which $\beta = 0$, our method gave positive evidence against the wrong model in 94% of the time. When the data were coming from the full model, our method gave very strong evidence against the wrong model in 83% of the time. In Figure 5, we drew the scatter plots for a few of these simulations.

2.3 Code used in R

Listing 6: The function bfac in R, used to compute BF_{01}, given the dataset and the prior parameters.

```
# bfac computes BF given the data set
# and the prior distributions

bfac<-function(datax, datay, c1=0.5,
    c2=0.5, p=0.01, q=0.01) {
mat<-numeric(1); n=length(datax);

Ev0=1/(q+0.5*sum(datay^2)-(sum(datay)^2)/(2*(n+c1)))^(p+n/2)

A=sqrt(c2+sum(datax^2)-(sum(datax)^2/(n+c1)))
C=sum(datax*datay)- (sum(datax)*sum(datay))/(n+c1)
D=c2+sum(datax^2)-(sum(datax)^2)/(n+c1)
B=(q+ 0.5*sum(datay^2)- 0.5*(sum(datay)^2)/(n+c1)
    -0.5*(C^2)*(1/D))^(p+n/2)
Ev1=sqrt(c2)*(1/A)*(1/B)
mat<-c(Ev0/Ev1)
return(mat) }
```

Listing 7: The code used in order to simulate data from the linear model $y = \alpha + \beta x + \varepsilon$, with and without the parameter β and compute the Bayes Factor BF_{01} for the two simulated data sets. It uses the function bfac as defined in Listing 6.

```
simulation<-function(n,datax=X, datay=Y, c1=0.5,
    c2=0.5, p=0.1, q=0.1) {
SimulationResults=matrix(ncol=2, nrow=n)
for (i in 1:n) {
    tau<-rgamma(1, shape=0.1, rate=0.1)
    epsilon<-rnorm(100, 0, sqrt(1/tau))
    alpha<-rnorm(1,0,sqrt(1/(c1*tau)))
    bbeta<-rnorm(1,0,sqrt(1/(c2*tau)))
    SimulationResults[i,1]=bfac(X,alpha+epsilon, p, q)
    SimulationResults[i,2]=bfac(X,alpha+bbeta*X+epsilon,p, q) }
return (SimulationResults) }
```

3 References

[ABCD] R. Agarwal, N. Barnett, P. Cerone, S. Dragomir, *A Survey on some Inequalitites for Expectation and Variance*, Computers and Mathematics with Applications, Vol. 49, pp. 429-480, 2005.

[CB01] G. Casella, R. Berger, *Statistical Inference*, Cengage Learning, 2001.

[KR95] R. Kass, A. Raftery, *Bayes Factors*, Journal of the American Statistical Association, Vol. 90, No. 430, pp. 773-795, 1995.

[Pet95] V. Petrov, *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*, Oxford Studies in Probability, Oxford University Press, 1995.