

Supplementary Material for: A novel algorithmic approach to Bayesian Logic Regression

A. HUBIN, G. STORVIK, F. FROMMLET

A GMJMCMC Algorithm

A.1 Tuning parameters

In all the simulations and in real data analysis we used the default tuning parameters of the implementation of MJMCMC downloaded from <http://aliaksah.github.io/EMJMCMC2016/>. The values which were used in the different simulation scenarios and for real data analysis for the parameters not related to MJMCMC but rather to the genetic algorithm part are presented in Table A.1.

Table A.1. Tuning parameters of GMJMCMC in the different examples (Ex.), where simple digits refer to the simulation scenario, RD1 refers to the Arabidopsis data analysis and RD2 to the Drosophila data analysis; Threads (Th.) - the number of CPUs utilized within the examples; N_{init} - the number of steps of MJMCMC during initialization; N_{expl} - the number of steps of MJMCMC between changes of population; M_{fin} - the number of unique models visited by MJMCMC for the final population; T_{max} - index of the final population; ρ_{min} - threshold for the trees to be deleted; P_{and} - probability of an *and* operator in crossovers and mutations; P_{not} - probability of using logical *not* in crossovers and mutations; P_c - probability of crossover to propose replacement trees; P_{init} - probability for a tree to be included into the initial solution for a new MJMCMC run in any iteration $t \geq 1$; ρ_{del} - probability of deletion in the reduction operator; C_{max} - maximal tree size allowed; k_{max} - maximal number of trees allowed in a model; d - size of population of genetic algorithm (number of trees searched by MJMCMC in each iteration).

Ex.	Th.	N_{init}	N_{expl}	M_{fin}	T_{max}	ρ_{min}	P_{and}	P_{not}	P_{init}	P_c	ρ_{del}	C_{max}	k_{max}	d
1	32	300	300	10000	16	0.2	1.0	0.2	0.5	0.9	0.5	5 (2)	10	15
2	32	300	300	10000	16	0.2	1.0	0.2	0.5	0.9	0.5	5 (2)	10	15
3	32	300	300	15000	33	0.2	0.9	0.1	0.5	0.9	0.5	5	10	15
4	32	300	300	10000	33	0.2	0.9	0.1	0.5	0.9	0.5	5	10	15
5	32	300	300	10000	33	0.2	0.9	0.1	0.5	0.9	0.5	5	10	20
6	32	250	250	20000	40	0.2	0.7	0.1	0.5	0.9	0.5	5	20	40
RD1	64	250	250	35000	40	0.2	0.7	0.1	0.5	0.9	0.5	5	15	25
RD2	64	250	250	15000	40	0.2	0.7	0.1	0.5	0.9	0.5	5	15	25

A.2 Theorem for parallel version of GMJMCMC

The following Theorem generalizes Theorem 1 from the manuscript to the parallelized version of GMJMCMC. Apart from letting the number of iterations go to infinity it is also possible to have only a finite number of iterations within each run but let the number of parallel runs go to infinity.

Theorem A.1. *Assume that we are running GMJMCMC in B parallel chains as describes in Section 2.3 of the manuscript. When the number of iterations within each chain b converges to infinity, the posterior estimates $\tilde{P}(\Delta | Y)$ of (16) from the manuscript will converge to $P(\Delta | Y)$.*

Assuming the search space \mathcal{S}_1 is selected randomly within the total set of possible search spaces and for a finite number of iterations within each chain b , the posterior estimates (16) will converge to $P(\Delta | Y)$ when $B \rightarrow \infty$.

Proof. When the number of iterations within each chain b converges to infinity, each $\tilde{P}_b(\Delta | Y)$ will converge to $P(\Delta | Y)$ according to Theorem 1 of the manuscript. Further, each $w_b \rightarrow 1/B$, proving the first part of the result.

When the initial search space \mathcal{S}_1 is selected randomly, any possible tree can be included. According to the construction of the initial model for the first MJMCMC run any model will have positive probability of being selected, giving the result directly. \square

Remark Selecting the search space \mathcal{S}_1 randomly among all possible models is in principle not easy due to the difficulty of specifying the complete model space. However, running the GMJMCMC algorithm with no data can be performed extremely fast, making it possible to select the initial population randomly.

B Details of Simulation Results

In this section we present further information on the simulation results of our six scenarios.

B.1 Binary Response

In case of GMJMCMC and FBLR a tree was counted as detected if its corresponding posterior probability was larger than 0.5. The power to detect a true tree is estimated by the percentage of simulation runs in which it was detected. The overall power is then defined as the average power over all individual true trees. A detected tree was counted as true positive if it was logically equivalent to a tree from the data generating model or to its logical complement, otherwise it was counted as false positive. FP denoted the average, over simulation runs, number of false positive detections and FDR was estimated as the average (over simulation runs) proportion of false discoveries, where this proportion was defined to be zero if there were no detections at all. WL is the number of binary covariates (leaves) which were not part of the data generating model but part of at least one detected tree.

Unfortunately, the output delivered by MCLR does not allow to compute the performance measures in the same way. Whenever MCLR detects a tree of size s then all subtrees are also reported as being detected. Furthermore MCLR reports for each detected tree only the set of leaves $v(L)$ and not the exact logical expression L itself. Thus it becomes impossible to define true positives by comparing the reported trees directly with the trees from the data generating model. Instead we considered for MCLR a reported tree L as a true positive whenever $v(L)$ coincided with the set of leaves of a true tree. This definition only gives an upper bound for the achieved power and is strongly biased in favor of MCLR. For the same reason, any reported tree that was a subtree of a true tree was not considered to be a false positive, resulting in only lower bounds of FP and FDR which are again strongly biased in favor of MCLR.

Table B.1 gives details about the frequencies of trees detected by the different methods. The first three lines give for each scenario the frequency with which the three true trees L_j in each scenario were

detected. All further detected trees are per definition false positives. In Scenario 1 and 2 there were occasionally trees of size 4 reported which were logically equivalent to the sum of two trees of size 2, for example $L_1^C \wedge L_2^C$ with a negative coefficient is equivalent to $L_1 + L_2$. These type of findings are specifically listed in Table B.1. For example the entry $85 + 6 + 8$ for Jeffrey's prior in Scenario 1 means that 85 times the data-generating tree per se was detected, 6 times the tree $L_1^C \wedge L_2^C$ and 8 times $L_1^C \wedge L_3^C$. The sum of these true positive findings is then used for the power reported in Table 1 of the manuscript.

With respect to false positives we considered different classes. The first class of false positives are trees which are comprised exclusively of leaves from a true tree L_j , typically subtrees or trees with a different logic expression. Based on the output of MCLR it is not possible to determine the frequencies of this kind of false positive detections as we will discuss in the next paragraph. In case of FBLR and GMJMCMC Table B.1 provides the frequency of this kind of trees in the rows labeled $v(L_j)$. For Scenario 1 and 2 we actually provide more detailed information. Here all true trees are of size 2 and almost all detected trees of the class $v(L_j)$ consisted of single leaves (the only exception was two instances of the expression $X_8^c \wedge X_{11}$ in Scenario 1 for FBLR). We therefore explicitly present the number of detections of the first leaf and the second leaf of L_j . The $v(M)$ rows give the number of trees combining leaves from different true trees. Finally the rows $WL(s)$ are concerned with the number of trees which include s leaves which were not in the data generating model at all.

In case of MCLR the same sort of classification is not possible due to the fact that MCLR does not report the exact logical tree L that it detects but only the corresponding set of leaves $v(L)$. Furthermore MCLR automatically reports the set of leaves for all subtrees of any detected tree which makes an assessment on how often these subtrees were actually detected by MCLR impossible. As a consequence we simply discarded reported subtrees when computing summary statistics, with one exception. In case of Scenario 3 MCLR reports 40 supertrees (trees for which a tree of interest is a subtree) of L_1 , which we classified as false positives themselves but which in principle play an important role for the determination of the power to detect L_1 . We ignored the fact that for any detected supertree of L_1 MCLR automatically also reports L_1 itself as detected and we pretended that in all these cases MCLR would actually have detected L_1 itself. Another peculiarity of MCLR is that it allows to search for trees of size 4, but that it does not report if it detected any such trees. In case of the four-way interaction L_3 from Scenario 3 there were 19 simulation runs where MCLR reported all four subtrees of size 3 from L_3 and we counted those instances as true positives, although MCLR did not really report the correct four-way interaction. For Scenario 1 and 2 none of these problems with supertrees occurred for MCLR because we restricted the search to trees of maximal size 2 in accordance with the data generating model.

The first two scenarios include only two-way interactions and we observe that GMJMCMC with Jeffrey's prior worked almost perfectly well, no matter whether using $C_{max=5}$ or $C_{max=2}$. In the few instances where it did not detect the correct tree it reported instead the two corresponding main effects, resulting for $C_{max} = 2$ in a total of 25 and 38 false positive, and for $C_{max} = 5$ in 8 and 3 false positive trees for the two scenarios (the corresponding average numbers of FP are also shown in Table 1 of the main manuscript). The robust g-prior resulted in a few more false positives which were mainly just single leaves instead of the two-way interactions. It is slightly surprising that in case of Jeffrey's prior GMJMCMC with $C_{max} = 5$ performed better than with $C_{max} = 2$. The reason for that is mainly that the trees of the form $L_i^C \wedge L_j^C$ were more likely to be selected than individual leaves. For the robust g prior this was apparently not always the case for L_3 .

FBLR chose in almost two thirds of the simulation runs two main effects instead of the correct interactions. The majority of the remaining false positives combined leaves from different true trees but there was also for each scenario one expression with a wrongly detected leaf, respectively. In contrast MCLR reported in approximately two thirds of the cases trees with the correct leaves resulting in larger power than for FBLR. On the other hand MCLR reported a much larger number of trees which combined leaves from different true trees than FBLR. MCLR reported only one tree with a wrong leaf in Scenario 2 and no such tree in Scenario 1. In summary we conclude that all three methods were doing extremely

Table B.1. Number of true and false positive trees for the three simulation scenarios with a binary response. A detailed description of the different classes of false positives ($v(L_j), v(M), WL(s)$) is given in the text above. The columns Jef. and Rob. g correspond to GMJMCMC with Jeffrey’s prior and with the Robust g-prior and the default setting $C_{max} = 5$, respectively. The columns Jef.(2) and Rob. g (2) give results for $C_{max} = 2$ for the sake of comparison with FBLR and MCLR.

	FBLR	MCLR	Jef.	Rob. g	Jef.(2)	Rob. g (2)
S.1						
L_1	30	67	85 + 6 + 8	90 + 3 + 7	97	98
L_2	42	61	89 + 6 + 4	89 + 7	100	95
L_3	33	59	83 + 8 + 4	50 + 3	91	77
$v(L_1)$	68+69	*	0	0	3 + 3	2 + 2
$v(L_2)$	54+53	*	0	3 + 3	1 + 0	5 + 5
$v(L_3)$	60+59(+2)	*	4 + 4	44 + 47 + 2	9 + 9	25 + 24
$v(M)$	22	270	0	2	0	0
WL(1)	1	0	0	0	0	0
S.2						
L_1	32	66	91 + 3 + 4	92 + 5 + 1	97	97
L_2	40	67	89 + 3 + 7	87 + 5 + 2	99	96
L_3	37	60	85 + 4 + 7	51 + 1 + 2	86	76
$v(L_1)$	64 + 66	*	1 + 1	2 + 2	3 + 3	3 + 3
$v(L_2)$	56 + 60	*	0	5 + 6 + 1	1 + 1	4 + 4
$v(L_3)$	56 + 56	*	4 + 4	44 + 46 + 2	15 + 15	26 + 26
$v(M)$	24	256	0	0	0	0
WL(1)	1	1	0	0	0	0
S.3						
L_1	93	93 (40SupT)	100	100		
L_2	4	67	91	56		
L_3	0	19 (SubT)	100	56		
$v(L_1)$	20	*	0	0		
$v(L_2)$	162	*	8	81		
$v(L_3)$	233	*	1	87		
$v(M)$	167	195	5	6		
WL(1)	34	54	1	0		
WL(2)	16	9	0	0		
WL(3)	8	0	0	0		

well in detecting the correct leaves in these simple scenarios but GMJMCMC was better than FBLR and MCLR in identifying the exact logical expressions.

The conclusion above is even more pronounced in the third scenario, which is more complex than the previous scenarios but still allows GMJMCMC with Jeffrey’s prior to perform almost perfectly. It detected both the two-way interaction L_1 and the four-way interaction L_3 with a power of 100%, and had only some minor difficulties to detect the three-way interaction L_2 . From the 15 false positive detections the majority consisted of subtrees of L_2 reported in those simulation runs where L_2 itself was not detected. Five trees were combinations of leaves from different true trees and there was only one tree including a leaf which was not part of the data generating model. GMJMCMC with the robust g-prior had substantially lower power to detect L_2 and L_3 but instead reported many corresponding subtrees. There were six reported logic expressions which mixed leaves from L_2 and L_3 . In comparison, both MCLR and FBLR performed much worse and only managed to detect L_1 with fairly large power. FBLR completely

failed to detect the higher order terms L_2 and L_3 whereas MCLR had at least some power to detect the three-way interaction L_2 . Both approaches reported way more false positive trees than GMJMCMC.

For FBLR we can discuss the structure of false positive detections in more detail. A large number of false positive expression were comprised of leaves from single true trees, 20 for $v(L_1)$, 162 for $v(L_2)$ and 233 for $v(L_3)$. These expressions were either subtrees of true trees or trees with misspecified logical operators and can be seen as substitutes for the true trees. Furthermore there were 167 logical expressions which combined leaves from different trees. Additionally, FBLR reported 34 trees with one wrongly detected leave, 16 trees with two wrongly detected leaves and even 8 trees of size three for which all leaves were not part of the data generating model. Thus apart from having problems with determining the exact form of the logical expressions in this scenario FBLR produced also a large number of false positive trees which have nothing to do with the correct model at all.

The performance of MCLR was only a little bit better. With respect to the results presented in Table 1 of the main manuscript it is now even more important than for the first two scenarios to emphasize that we are dealing with upper bounds of the power and lower bounds of the number of false positives. MCLR automatically reports all subtrees of any detected tree which makes an assessment how often these tree were actually detected by MCLR impossible. As a consequence we simply discarded reported subtrees from further statistical analysis, with one exception. In case of Scenario 3 MCLR reports some supertrees of L_1 , which we classified as false positives themselves but which in principle played an important role for the determination of the power of L_1 . We ignored the fact that for any detected supertree of L_1 MCLR automatically also reports L_1 itself as detected and pretend that in all these cases MCLR would actually have detected L_1 itself. Another peculiarity of MCLR is that it allows to search for trees of size 4, but that it does not report if it detected any such trees. In case of the four-way interaction L_3 there were 19 simulation runs where MCLR reported all four subtrees of size 3 from L_4 and we counted those instances as true positives, although MCLR did not really report the correct four-way interaction. For Scenario 1 and 2 none of these problems with supertrees occurred for MCLR because we restricted the search to trees of maximal size 2 in accordance with the data generating trees.

Not counting any subtrees of reported trees as false positives gives MCLR a huge advantage, nevertheless it reported almost 20 times more false positive expressions than GMJMCMC. Among those were 40 supertrees of L_1 , which all contributed to the power of L_1 although it is not guaranteed that in all corresponding simulation runs L_1 itself was actually detected. There were 195 false positive trees which combined leaves from different true trees. It was more problematic that there were 54 trees with one wrongly detected leave and 9 trees with two wrongly detected leaves. While there were not as many trees which were completely wrong as for FBLR there were still a considerable number of leaves reported by MCLR which were not part of the data generating model.

B.2 Continuous Response

Table B.2 gives detailed results about the frequencies of detected trees similarly to Table B.1 but now for the three linear regression scenarios. At the beginning we have again for each scenario the number of true positives with L_j referring to the trees of the data generating model. As described above we split the detections of false positives again in the classes $v(L_j)$ which refers to logic expressions consisting only of leaves from L_j , $v(M)$ which refers to logic expressions consisting of leaves from the data generating model but mixing leaves from different trees, and $WL(1)$ corresponding to trees including one wrong leave. For the last expression of Scenario 6 it holds that $L_8 = X_{11} \wedge X_{13} + X_{19} \wedge X_{50} - X_{11} \wedge X_{13} \wedge X_{19} \wedge X_{50}$ and it turned out that in many simulation runs GMJMCMC was detecting the three expression from the alternative version. In the main manuscript we considered these findings potentially as true positives and in Table B.2 we explicitly report the frequency of detection for each of these trees.

There is not much to be said about Scenario 4 apart from the fact that the only false positive detection $L_2 \vee L_3$ was very close to the expression $L_2 + L_3$ of the data generating model. In Scenario 5 the results using Jeffrey's prior and the robust g-prior are very similar. For those trees which were detected in all

Table B.2. Detailed results for the three simulation scenarios for linear regression. A detailed description of the different classes of false positives ($v(L_j)$, $v(M)$, $WL(1)$) is given in Section B.1. The columns Jef. and Rob. g correspond to GMJMCMC with Jeffrey’s prior and with the Robust g-prior, respectively. In Scenario 4 there was only one false positive detection which is listed explicitly. In Scenario 6 frequencies of the three trees which in combination give L_8 are also listed explicitly.

Scenario 4	Jef.	Rob.g	Scenario 6	Jef.	Rob.g
L_1	100	100	L_1	95	99
L_2	99	100	L_2	98	99
L_3	97	98	L_3	98	99
$L_2 \vee L_3$	1		L_4	96	95
Scenario 5			L_5	100	100
L_1	100	100	L_6	95	96
L_2	100	99	L_7	32	45
L_3	96	100	L_8	21	16
L_4	89	90	$X_{11} \wedge X_{13}$	76	78
$v(L_2)$	0	2	$X_{19} \wedge X_{50}$	75	81
$v(L_3)$	12	0	$X_{11} \wedge X_{13} \wedge X_{19} \wedge X_{50}$	72	69
$v(L_4)$	22	19	$v(L_3)$	6	2
$v(M)$	1	2	$v(L_4)$	12	15
$WL(1)$	2	5	$v(L_5)$	0	2
			$v(L_6)$	18	13
			$v(L_7)$	84	70
			$v(L_8)$	24	49
			$v(M)$	58	38
			$WL(1)$	3	7

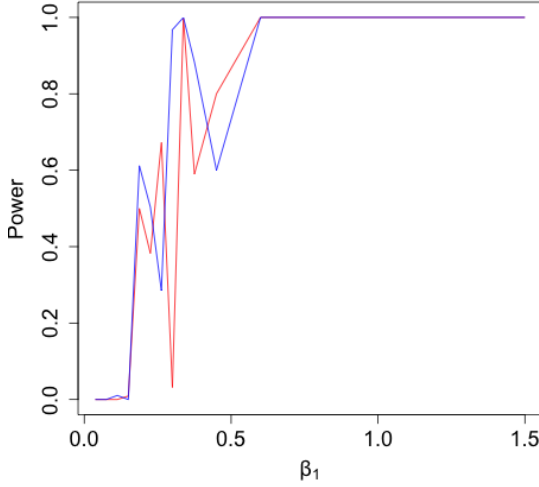
simulation runs (L_1 and L_2 for Jeffrey’s, L_1 and L_3 for the robust g-prior) no false positive subtrees were reported. The majority of false positives for both priors is comprised of subtrees and there are only a very small number of detections which combine leaves from two different true trees (1 for Jeffrey’s and 2 for the robust g-prior). Finally GMJMCMC with Jeffrey’s prior reported two trees of size 4 and size 5, respectively, each of which included the wrongly detected leave X_{43} , whereas GMJMCMC with the robust g-prior reported five trees which included wrong leaves.

For the most complex Scenario 6 once again Jeffrey’s prior and the robust g-prior perform quite similar. For the first 6 data generating trees both priors have very large power. For L_7 the power is much lower and both priors report a large number of subtrees which are counted as false positives. For L_8 the alternative representation of the logic expression has been discussed in the main manuscript. Only 58 false positive trees for Jeffrey’s prior and 38 for the robust g-prior combined leaves from different true trees. The number of trees including one wrongly detected leave was 3 and 7, respectively, which once more illustrates that GMJMCMC is very good at controlling the type I error when it comes to including leaves which have nothing to do with the data generating model.

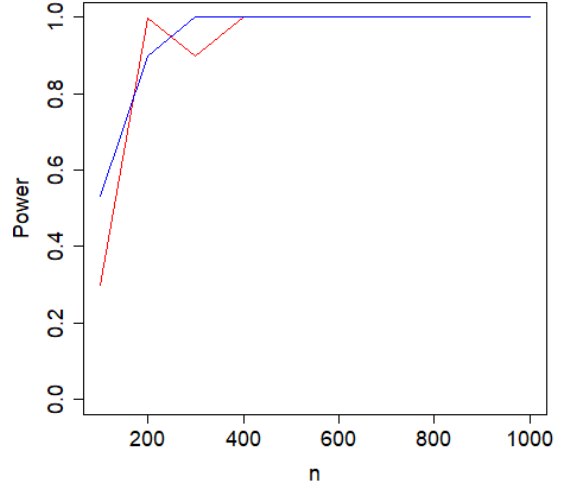
C Sensitivity Analysis

Figures S1 - S3 present the sensitivity results for L_1 , L_2 and L_3 , respectively, similarly to Figure 1 from the main text (which is concerned with the change of power in detecting L_4 when certain parameters are varied).

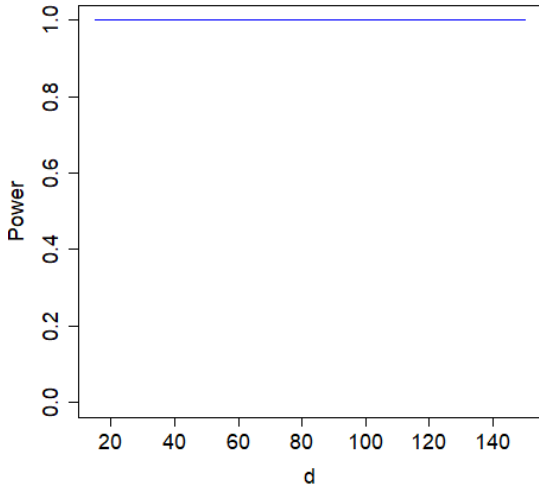
1) Regression coefficient: β_1



2) Sample size: n



3) Population size: d



4) Correlation of misspecified leave: r

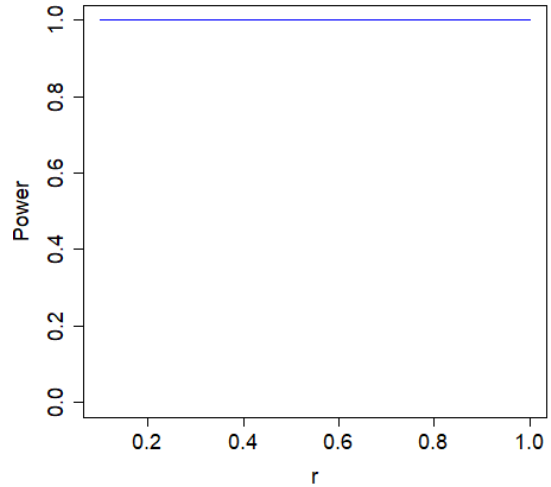


Figure S1. Dependence of power to detect L_1 for Jeffrey's prior (red) and the robust g-prior (blue) when varying different parameters as specified above each plot. Parameters which are not explicitly varied are kept fixed at the levels from the original Scenario 5, except for the first plot where all four coefficients $\beta_1 \dots, \beta_4$ are simultaneously varied by multiplying with the same factor.

It is not surprising that changing the correlation of the misspecified leave in L_4 has hardly any effect on the power of $L_1 - L_3$. The slight drop observed for L_3 when using the robust g-prior for fairly large values of r appears to be due to sampling variation and the small number of repetitions we have used. Changing the population size d did not influence the power for L_1 and L_3 . Only L_2 was not detected once (Jeffrey) or twice (robust g)) for the smallest value of d .

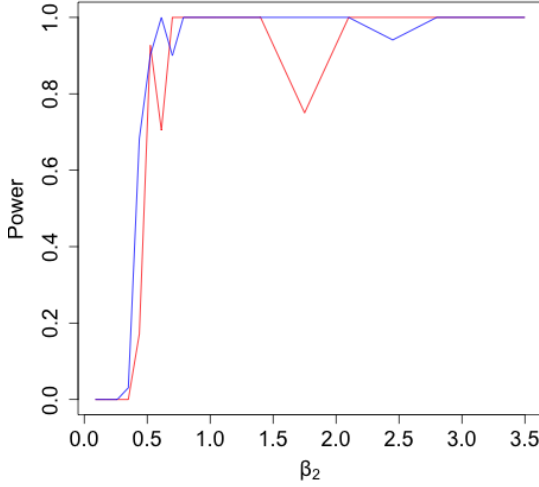
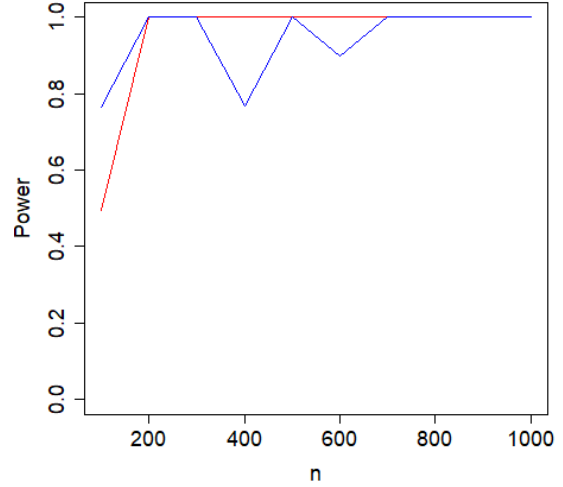
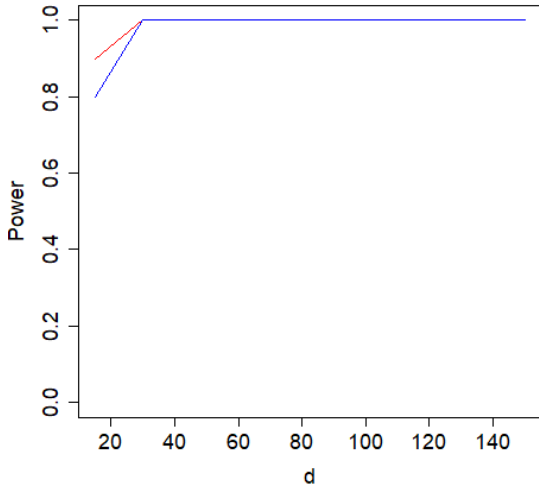
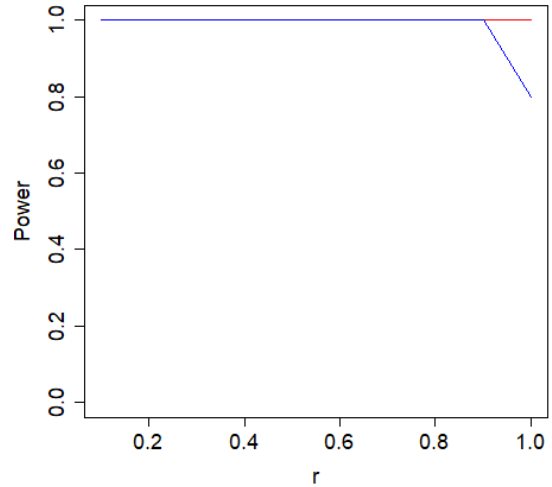
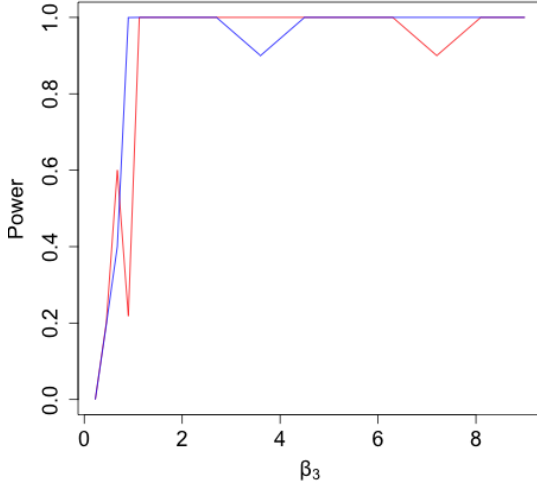
1) Regression coefficient: β_2 2) Sample size: n 3) Population size: d 4) Correlation of misspecified leave: r 

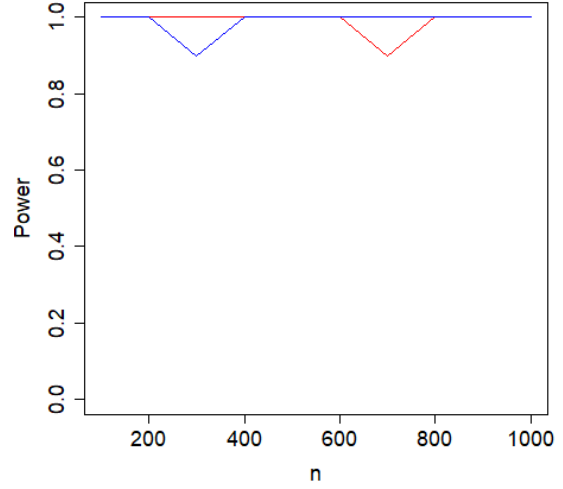
Figure S2. Dependence of power to detect L_2 for Jeffrey's prior (red) and the robust g-prior (blue) when varying different parameters as specified above each plot. Parameters which are not explicitly varied are kept fixed at the levels from the original Scenario 5, except for the first plot where all four coefficients $\beta_1 \dots, \beta_4$ are simultaneously varied by multiplying with the same factor.

The observed drop in power depending on the coefficients and the sample size was as expected for L_1 and L_2 . It is note-worthy that for L_3 the power starts to drop only for coefficients smaller than 2. The original value $\beta_3 = 9$ is so large that even with a sample of size 100 the estimated power still remains close to 1.

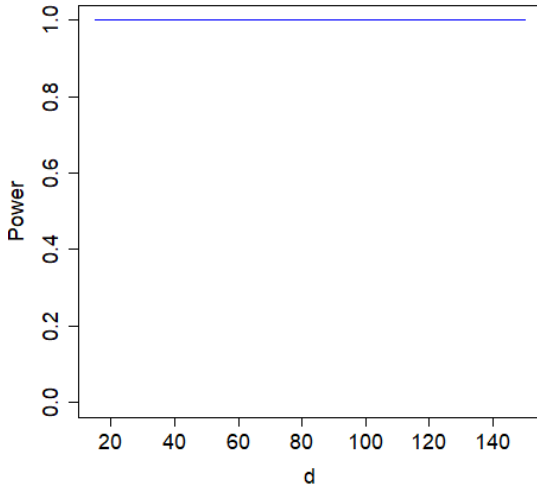
1) Regression coefficient: β_3



2) Sample size: n



3) Population size: d



4) Correlation of misspecified leave: r

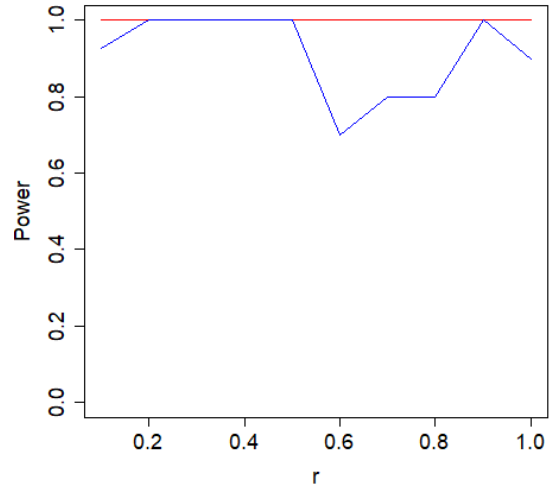


Figure S3. Dependence of power to detect L_3 for Jeffrey's prior (red) and the robust g-prior (blue) when varying different parameters as specified above each plot. Parameters which are not explicitly varied are kept fixed at the levels from the original Scenario 5, except for the first plot where all four coefficients $\beta_1 \dots, \beta_4$ are simultaneously varied by multiplying with the same factor.

For the z-test it holds that increasing the effect size by a factor K has the same effect on the power like increasing the sample size by a factor K^2 . In linear regression models a similar relationship holds approximately with respect to increasing regression coefficients. Figure S4 illustrates that this kind of relationship seems to hold also in the case of our Bayesian logic regression approach. The power curves for varying β coefficients with a sample size $n = 1000$ appear to be more or less the same than power curves with coefficients multiplied by a factor $1/\sqrt{10}$ and the sample size multiplied by a factor 10.

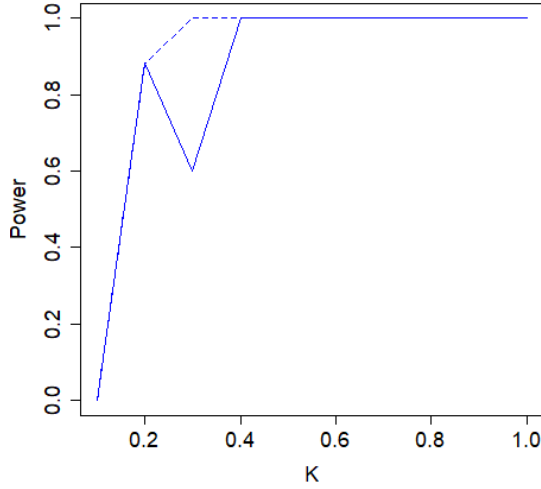
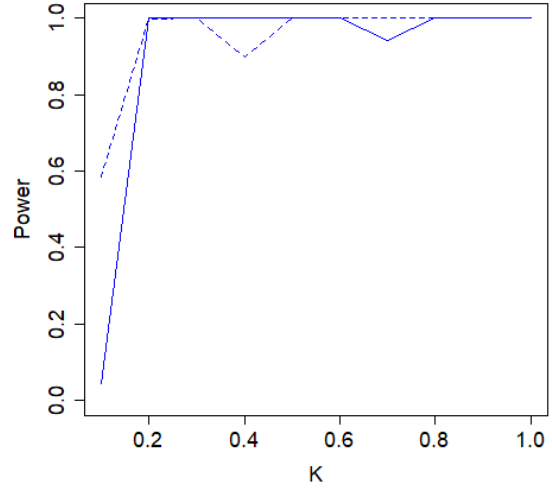
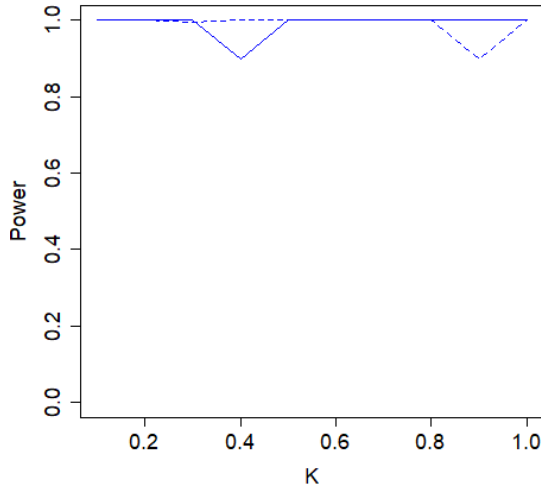
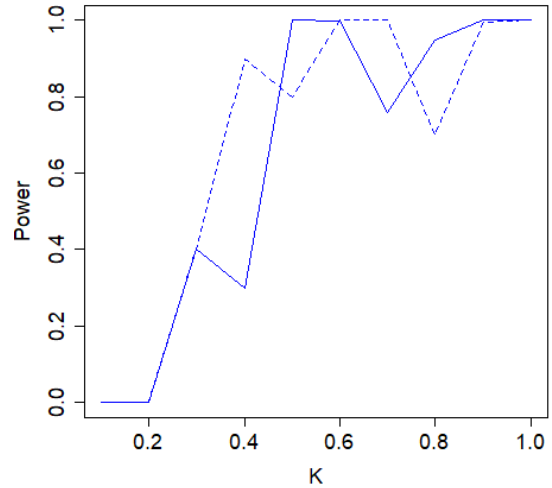
Power for L_1 Power for L_2 Power for L_3 Power for L_4 

Figure S4. Power to detect L_1 - L_4 with regression coefficients $(K\beta_1, K\beta_2, K\beta_3, K\beta_4)$ with sample size $n = 1000$ (solid) and with $(K\beta_1^*, K\beta_2^*, K\beta_3^*, K\beta_4^*)$ with sample size $n = 10000$ (dashed) where $\beta_j^* = \beta_j/\sqrt{10}$. Results are only shown for the robust g-prior.

D Real Data Analysis on *Drosophila*

Drosophila

As a second real data example we considered the *Drosophila* back cross data from Zeng et al. (2000)¹. There are five quantitative traits available for each species (abbreviated as **pc1**, **adjpc1**, **area**, **areat** and **tibia**) which quantify the size and shape of the posterior lobe of the male genital arch. The original publication (Zeng et al., 2000) only includes results on the first measure **pc1**, which was later analyzed for epistatic effects using a model selection approach based on the Cockerham coding (Bogdan et al., 2008).

Compared with the Arabidopsis example this backcross data set has a much larger sample size combined with a smaller number of genetic markers, which both helps to increase the power to detect QTL. Genotype data from 45 markers is available for 471 samples from *Drosophila Simulans* and 491 samples from *Drosophila Mauritana*. Six markers are located on chromosome X, 16 markers on chromosome 2 and 23 markers on chromosome 3. Imputation of the few missing genotypes was performed by a simple maximum likelihood approach based on flanking markers. More details on the experiments and the measured traits can be found in Zeng et al. (2000).

Marker	Chr	Marker name	$\tilde{P}(L Y)$	mBIC	Signif.
m2	X	w	1.000	x	***
m4	X	v	1.000	x	***
m7	2	gl	0.960	x	***
m9	2	cg	1.000		***
m10	2	gpdh		x	
m14	2	mhc	1.000	x	***
m18	2	sli	0.414	x	**
m22	2	zip	0.838	x	***
m23	2	lsp	0.998	x	***
m26	3	dbi	1.000	x	***
m29	3	fz	1.000	x	***
m32	3	rdg		x	
m33	3	ht	1.000		***
m35	3	ninaE		x	
m37	3	mst	1.000	x	***
m40	3	hb	0.942		
m41	3	rox		x	
m44	3	jan	1.000	x	***
m12, m34	2, 3	glt \wedge ant		x	
m11, m35	2, 3	ninaE \wedge ninaC	0.998		***

Table D.1. Results for *Drosophila Simulans* are presented for the trait **pc1** from Zeng et al. (2000). Posterior probabilities for additive and epistatic effects detected with GMJMCMC (column $\tilde{P}(L | Y)$) are compared with the findings reported by Bogdan et al. (2008) using mBIC as a selection criterion (column mBIC). The last column gives the level of confidence, similar to Table 3 from the main manuscript.

Table D.1 reports trees with posterior probabilities larger than 0.3 for the trait **pc1** of *Drosophila Simulans* and compares with the model obtained with mBIC - based forward selection by Bogdan et al. (2008). The logic regression approach detected most of the main effects also previously reported, which

¹Data downloaded from <ftp://statgen.ncsu.edu/pub/qlcart/data/zengetal99>. There one can also find a linkage map in centiMorgan for the markers on three different chromosomes

Marker	Chr	Marker name	$\hat{P}(L Y)$	mBIC	Signif.
m1	X	ewg		x	
m4	X	v	0.994	x	***
m9	2	cg	1.000	x	***
m11	2	ninaC	0.382	x	**
m15	2	ddc	1.000	x	***
m18	2	sli	0.523	x	***
m22	2	zip	1.000	x	***
m24	3	ve	0.966		***
m25	3	acr		x	
m26	3	dbi	0.995		
m28	3	cyc	0.398	x	**
m29	3	fz	0.834		
m34	3	ant	1.000	x	***
m37	3	mst		x	
m39	3	tub	0.999		***
m40	3	hb		x	
m41	3	rox	0.420		**
m44	3	jan	1.000	x	***
m1, m2	X, X	w\ewg	0.855		***
m2, m36	X, 3	w\fas		x	
m29, m40	3, 3	fz\hb		x	

Table D.2. Results for *Drosophila Mauritana* are presented for the trait **pc1** from Zeng et al. (2000). Posterior probabilities for additive and epistatic effects detected with GMJMCMC (column $\hat{P}(L | Y)$) are compared with the findings reported by Bogdan et al. (2008) using mBIC as a selection criterion (column mBIC). Posterior probabilities are only reported for trees with $\hat{P}(L | Y) > 0.3$ are reported. The last column gives the level of confidence, similar to Table 3 from the main manuscript.

in itself is quite interesting because as we allowed for higher order interactions we looked at a much larger model space and used therefore implicitly larger penalties than mBIC. In two locations GMJMCMC preferred a neighboring marker (*cg* instead of *gpdh* on chromosome 2 and *hb* instead of *rox* on chromosome 3. In one region on chromosome 3 mBIC selected 2 markers (*rdg*, *ninaE*) whereas GMJMCMC selected only one marker in the middle. These kind of discrepancies are quite natural due to marker correlations in back cross data (Bogdan et al., 2008). Just like with the mBIC approach we detected a two-way interaction between chromosome 2 and chromosome 3, where on both locations the two methods chose neighboring markers, respectively. Otherwise the epistatic effect detected with both methods is identical.

Table D.2 contains the corresponding results for *Drosophila Mauritana*. As before GMJMCMC detects most of the additive effects that were reported by mBIC, though it sometimes chooses flanking markers (*ve* and *dbi* instead of *acr*, *tub* instead of *hb*). Interestingly the marker *ewg* on the X-chromosome is not reported as a main effect but rather as a two-way interaction together with *v* also on the X-chromosome, which also shows up as an additive effect. On the other hand the two-way interactions obtained with mBIC are not confirmed. Instead of the interaction between *fz* and *hb* GMJMCMC reports additional main effects on *fz* and *rox* (the neighbor of *hb*). For the interaction between *w* and *fas* there are no substitutes detected.

The results for the other four traits (**adjpc1**, **area**, **areat** and **tibia**) which were neither analyzed by Zeng et al. (2000) nor by Bogdan et al. (2008) are provided in the following two tables. In case of *Drosophila Simulans* we detected three two-way interactions for *adjpc1*. For *Drosophila Mauritana* further two-way interactions were found; two for **adjpc1**, three for **area**, and two more for **areat**. We did

Table D.3. Posterior probabilities for additive and epistatic effects detected with GMJMCMC for four additional traits: *Drosophila simulans*.

Population	Phenotype	Chr	Marker name	$\tilde{P}(L Y) > 0.5$
Simulans	adjpc1	3	rox	1.000
Simulans	adjpc1	3	dbi	1.000
Simulans	adjpc1	2	gpdh	1.000
Simulans	adjpc1	X	v	1.000
Simulans	adjpc1	2	plu	1.000
Simulans	adjpc1	3	mst	0.999
Simulans	adjpc1	2^3	(gl)^(fz)	0.998
Simulans	adjpc1	3	efi	0.985
Simulans	adjpc1	X	w	0.984
Simulans	adjpc1	3	fz	0.983
Simulans	adjpc1	3^3	(lsp)^(ht)	0.982
Simulans	adjpc1	2^3	(duc)^(fas)	0.978
Simulans	area	3	fz	1.000
Simulans	area	2	mhc	1.000
Simulans	area	3	jan	1.000
Simulans	area	X	w	1.000
Simulans	area	3	dbi	1.000
Simulans	area	X	v	0.999
Simulans	area	3	rox	0.998
Simulans	area	3	ninaE	0.996
Simulans	area	3	ve	0.990
Simulans	area	2	ninaC	0.970
Simulans	area	2	zip	0.952
Simulans	area	3	ht	0.864
Simulans	area	2	cg	0.806
Simulans	areat	3	jan	1.000
Simulans	areat	2	mhc	1.000
Simulans	areat	X	w	1.000
Simulans	areat	3	tub	1.000
Simulans	areat	3	rox	1.000
Simulans	areat	3	ninaE	1.000
Simulans	areat	3	fz	1.000
Simulans	areat	X	v	1.000
Simulans	areat	3	dbi	1.000
Simulans	areat	2	ninaC	1.000
Simulans	areat	2	zip	1.000
Simulans	areat	3	ve	1.000
Simulans	areat	3	ht	0.952
Simulans	areat	2	cg	0.925
Simulans	tibia	X	run	0.747

not find higher order interactions for any of these traits and based on the experience from our simulation study we might conclude that there are actually at least no strong higher order epistatic effects.

Table D.4. Posterior probabilities for additive and epistatic effects detected with GMJMCMC for four additional traits: *Drosophila Mauritiana*.

Population	Phenotype	Chr	Marker name	$\tilde{P}(L Y) > 0.5$
Mauritiana	adjpc1	2	cg	1.000
Mauritiana	adjpc1	3	ant	1.000
Mauritiana	adjpc1	3	jan	1.000
Mauritiana	adjpc1	3	acr	1.000
Mauritiana	adjpc1	3	eip	1.000
Mauritiana	adjpc1	3 \vee 3	(cyc) \vee (hb)	1.000
Mauritiana	adjpc1	2	gl	1.000
Mauritiana	adjpc1	2	sli	1.000
Mauritiana	adjpc1	X	ewg	1.000
Mauritiana	adjpc1	3 \vee X	(mst) \vee (v)	0.999
Mauritiana	area	3	ant	1.000
Mauritiana	area	3	jan	1.000
Mauritiana	area	2	cg	1.000
Mauritiana	area	2	zip	0.999
Mauritiana	area	3	acr	0.990
Mauritiana	area	3	fz	0.985
Mauritiana	area	3 \vee X	(ve) \vee (w)	0.984
Mauritiana	area	X	ewg	0.958
Mauritiana	area	3 \vee X	(tub) \vee (v)	0.890
Mauritiana	area	3	rox	0.873
Mauritiana	area	3 \vee 3	(cyc) \vee (tub)	0.862
Mauritiana	area	2	sli	0.714
Mauritiana	area	2	ninaC	0.613
Mauritiana	area	2	mhc	0.535
Mauritiana	areat	3	ant	1.000
Mauritiana	areat	3	efi	1.000
Mauritiana	areat	2	zip	1.000
Mauritiana	areat	2	cg	1.000
Mauritiana	areat	X	ewg	1.000
Mauritiana	areat	3	rox	1.000
Mauritiana	areat	X \vee 3	(v) \vee (mst)	1.000
Mauritiana	areat	3	fz	0.996
Mauritiana	areat	3 \wedge 2	(1-(fz)) \wedge (ninaC)	0.974
Mauritiana	areat	3	acr	0.973
Mauritiana	areat	3	cyc	0.685
Mauritiana	tibia	X	v	0.999
Mauritiana	tibia	3	hb	0.999
Mauritiana	tibia	2	mhc	0.997
Mauritiana	tibia	2	plu	0.625

References

- Bogdan, M., Frommlet, F., Biecek, P., Cheng, R., Ghosh, J. K., and Doerge, R. W. (2008). “Extending the Modified Bayesian Information Criterion (mBIC) to Dense Markers and Multiple Interval Mapping.” *Biometrics*, 64(4): 1162–1169.
- Zeng, Z. B., Liu, J., Stam, L. F., Kao, C. H., Mercer, J. M., and Laurie, C. C. (2000). “Genetic architecture of a morphological shape difference between two *Drosophila* species.” *Genetics*, 154: 299–310.