

SAT-SAMPLING FOR STATISTICAL SIGNIFICANCE TESTING IN SPARSE CONTINGENCY TABLES

PATRICK SCHARPFENECKER* AND TOBIAS WINDISCH

ABSTRACT. Exact conditional tests for contingency tables require sampling from fibers with fixed margins. Classical Markov basis MCMC is general but often impractical: computing full Markov bases that connect all fibers of a given constraint matrix can be infeasible and the resulting chains may converge slowly, especially in sparse settings or in presence of structural zeros. We introduce a SAT-based alternative that encodes fibers as Boolean circuits which allows modern SAT samplers to generate tables randomly. We analyze the sampling bias that SAT samplers may introduce, provide diagnostics, and propose practical mitigation. We propose hybrid MCMC schemes that combine SAT proposals with local moves to ensure correct stationary distributions which do not necessarily require connectivity via local moves which is particularly beneficial in presence of structural zeros. Across benchmarks, including small and involved tables with many structural zeros where pure Markov-basis methods underperform, our methods deliver reliable conditional p-values and often outperform samplers that rely on precomputed Markov bases.

1. INTRODUCTION

Exact conditional inference for discrete exponential families, such as independence testing in contingency tables, hinges on computations over *fibers*—the sets of nonnegative integer tables with fixed sufficient statistics. A standard approach is to construct an irreducible Markov chain over a fiber using a Markov basis and to approximate conditional p -values by Markov chain Monte Carlo (MCMC). The use of Markov-basis-style moves and exact conditional logic has recently extended to domains such as neural spatial interaction models Zachos et al. (2024), underscoring broader relevance of scalable sampling over constrained integer lattices.

While widely applicable in theory, this strategy faces two well-known bottlenecks in practice. First, computing a full Markov basis *a priori* is often computationally prohibitive or outright infeasible for realistic models and table sizes Aoki and Takemura (2003). Second, even when a basis is available, the induced moves can mix slowly and require substantial tuning effort Windisch (2016). These issues are exacerbated by structural zeros and other constraints, which can inflate the size of Markov bases and complicate connectivity Rapallo (2006). To mitigate these challenges, several lines of work have aimed to reduce—and sometimes eliminate—the dependence on precomputed Markov bases. Dynamic Markov bases

E-mail addresses: {patrick.scharpfenecker,tobias.windisch}@hs-kempten.de.

Both authors are with the University of Applied Sciences Kempten, Germany.

* Corresponding author.



FIGURE 1. Fiber walks using Markov-basis on left-hand side and SAT-sampler proposals (in red) combined with Markov moves (in blue) on right-hand side.

compute moves on the fly, tailored to the current region of the fiber Dobra (2011). Alternative proposals leverage lattice-basis moves and study when such bases suffice for connectivity Hazelton and Karimi (2024), or develop samplers that forgo Markov bases entirely in favor of more direct lattice-based proposals Hara et al. (2011). More recently, RUMBA introduced randomized updates that sample lattice points efficiently in high dimensions Bakenhus and Petrović (2024). In parallel, problem-specific strategies have been proposed for independence testing in large and sparse tables Zhang (2019). Collectively, these methods trade off generality, ease of implementation, and statistical efficiency, and they highlight a core need: flexible, scalable samplers that deliver reliable conditional p -values without incurring the full cost of Markov-basis computation.

We explore a different approach by encoding fibers as solutions to constraint satisfaction problems and solve them using SAT and SMT solvers. Despite SAT’s NP-completeness, modern conflict-driven clause-learning (CDCL) solvers routinely handle large, structured instances with remarkable effectiveness. Representative systems include CryptoMiniSat5, a Minisat-derived solver extended with native XOR reasoning for cryptographic constraints Soos (2016); Sörensson (2010), and more recent high-performance solver such as Kissat Biere et al. (2024). Recently, there were advances in *sampling* from SAT instances (e.g. UniGen3 Soos et al. (2020), CMSGen Golia et al. (2021)) and from SMT instances (e.g. SMT-Sampler Dutra et al. (2018a), GuidedSampler Dutra et al. (2019) and MeGASampler MeG (2025)), allowing faster and (almost-)uniform sampling from problem instances. While SMT samplers claim to be faster in most instances than using bit-blasting and then SAT samplers, especially the most recent and for our problem applicable sampler GuidedSampler, our initial experiments seemed to indicate that this might not be the case for our problem. Performance of these competing approaches depend on the concrete problem instances. See e.g. Shaw and Meel (2024), another recent result which gets better results on their problem instances using bit-blasting and SAT sampling over SMT sampling. We therefore focus on bit-blasting and applying SAT samplers. These SAT-based tools have found applications in software testing, configurable systems, and bug synthesis Dutra et al. (2018b); Heradio et al. (2022); Xiang et al. (2022); Roy et al. (2018).

Our premise is that fibers underlying exact conditional tests can be encoded compactly as Boolean circuits and then translated to conjunctive normal form (CNF) via a Tseitin-style transformation. This encoding captures the linear margin constraints, structural zeros, and bound constraints that define the fiber. Once in CNF, SAT samplers can be used to

draw tables from the fiber; enumerating SAT solutions and applying extraction routines can enumerate tables exactly for small instances; and CDCL machinery can act as a highly optimized engine for constraint propagation and randomized exploration.

However, integrating SAT-based sampling methods into the fiber-walk framework for exact conditional inference requires a careful design as SAT samplers typically do not yield perfectly uniform samples and the transition probabilities between fiber elements are not explicitly known. We address these issues by only selectively incorporating SAT proposals into hybrid MCMC schemes and still maintaining the correct stationary distribution after applying Metropolis-Hastings corrections (see Figure 1). Our SAT-viewpoint offers new advantages as it not only bypasses the need to compute a global Markov basis beforehand but also yields mixing benefits particularly in settings where the combinatorial structure of the fibers is complex like in the presence of structural zeros or small right hand sides.

This work is organized as follows. After reviewing log-linear models and fiber graphs in Section 2, we provide an efficient encoding of fiber constraints into Boolean circuits and a Tseitin translation to CNF that preserves sparsity and enables strong unit propagation in CDCL solvers in Section 3. In Section 4, we quantify the extent and structure of sampling biases in state-of-the-art SAT samplers and develop practical mitigation techniques that improve the accuracy of conditional p -values without sacrificing correctness. Finally, in Section 5, we present empirical results from a comprehensive benchmark of fiber-walk strategies for contingency tables with and without structural zeros demonstrating that SAT-based methods can outperform traditional Markov-basis approaches in terms of mixing speed.

2. LOG-LINEAR MODELS AND FIBER GRAPHS

Let $\Omega = \{\omega_1, \dots, \omega_d\}$ and let $\mathcal{P}_\Omega$ the set of all probability mass functions $\pi : \Omega \rightarrow [0, 1]$ on Ω , i.e. $\sum_{i=1}^d \pi(\omega_i) = 1$. To ease notation, we will associate the probability mass functions on Ω with vectors $\pi \in [0, 1]^d$ having $\|\pi\|_1 = 1$ by setting $\pi_i := \pi(\omega_i)$. The log-linear model associated to $A \in \mathbb{Z}^{k \times d}$ is the set of distributions on Ω :

$$\mathcal{P}_A := \{\pi \in \mathcal{P}_\Omega : (\log \pi_1, \dots, \log \pi_d) \in \text{rowspan}(A)\}.$$

A typical question in statistical inference is whether iid samples $x_1, \dots, x_n \in \Omega$ of an unknown probability mass function $\bar{\pi} \in \mathcal{P}_\Omega$ provide evidence for or against $\bar{\pi} \in \mathcal{P}_A$. Log-linear models constitute a broad and general class of models including, for instance, hierarchical models. A prominent log-linear model is the *independence model* Drton et al. (2009), where two jointly distributed categorical variables should be tested on statistical independence, as well as the *no-three way interaction model* (Drton et al., 2009, Section 1.2). Typically, the frequency counts $u^{\text{obs}} \in \mathbb{N}^d$ with $u_i^{\text{obs}} := |\{j \in [n] : x_j = \omega_i\}|$ for $i \in [d]$ of the observations is the central object of the analysis, which is presented in many log-linear models as a multi-dimensional *contingency table*. One can think of a contingency table as a histogram of the observed data. Statistical significance tests become more involved in degenerated cases, like if the tables are *incomplete*, that is if certain entries $S \subset [d]$ called *structural zeros* are impossible and thus $\pi_i = 0$ for all $i \in S$. Typically, structural zeros can be included into the model by unit vectors e_i , $i \in S$ as rows (see also Section 5.1). Now, to test $\bar{\pi} \in \mathcal{P}_A$, a first step is to compute a *maximum-likelihood estimator* $\tilde{\pi}$ from $\mathcal{P}_A$ for the samples $x_1, \dots, x_n \in \Omega$, that is an element of (the possibly empty set) $\arg \max_{\pi \in \mathcal{P}_A} \prod_{i=1}^n \pi(x_i)$. Let $b := A \cdot u^{\text{obs}}$ be the *marginals*, then

the set of all frequency counts having the same marginals under A is called the b -fiber of A : $\mathcal{F}_{A,b} = \{u \in \mathbb{N}^d : A \cdot u = b\}$. An *extremeness-measure* $X : \mathbb{N}^d \rightarrow \mathbb{R}$ quantifies how extreme a given frequency count $u \in \mathbb{N}^d$ with $\|u\|_1 = n$ is compared to the frequencies obtained by the maximum likelihood estimator (see Algorithm 1 for a popular choice). Using X and defining

$$(2.1) \quad \rho : \mathcal{F}_{A,b} \rightarrow [0, 1], \rho(u) \sim \frac{1}{v_1! \cdots v_d!}$$

and $f : \mathcal{F}_{A,b} \rightarrow \{0, 1\}$ with $f(v) = \mathbf{1}_{X(v) \geq X(u^{\text{obs}})}$ the value $\mathbb{E}_\rho[f] = \sum_{u \in \mathcal{F}_{A,b}} f(u) \rho(u)$ constitutes the *conditional p-value* that the unknown distribution π yielding the frequency counts u^{obs} is from $\mathcal{P}_A$. Clearly, the normalizing constant of ρ together with the size of $\mathcal{F}_{A,b}$ make it hard to compute $\mathbb{E}_\rho[f]$ directly and thus sampling from $\mathcal{F}_{A,b}$ is required to approximate it.

A classic way to approximate conditional p -values is to construct a Markov chain that converges against ρ on $\mathcal{F}_{A,b}$ using the *Metropolis-Hastings* methodology Metropolis et al. (1953); Hastings (1970). Generally, given a graph G on $\mathcal{F}_{A,b}$ and a random walk $\mathcal{W}$, the *Metropolis-Hastings walk* $\mathcal{M}_{\mathcal{W},\rho}$ is the random walk on G defined by

$$\mathcal{M}_{\mathcal{W},\rho}(u, v) := \begin{cases} \mathcal{W}(u, v) \cdot \min \left\{ 1, \frac{\rho(v)\mathcal{W}(v,u)}{\rho(u)\mathcal{W}(u,v)} \right\}, & \text{if } u \neq v \\ 1 - \sum_{w \in \mathcal{F}_{A,b} \setminus \{u\}} \mathcal{W}(u, w) \cdot \min \left\{ 1, \frac{\rho(w)\mathcal{W}(w,u)}{\rho(u)\mathcal{W}(u,w)} \right\}, & \text{if } u = v \end{cases}.$$

Many properties of the random walk $\mathcal{W}$ hold also for $\mathcal{M}_{\mathcal{W},\rho}$. Particularly, if $\mathcal{W}$ is irreducible, aperiodic, and if $\mathcal{W}(u, v) > 0$ if and only if $\mathcal{W}(v, u) > 0$ for all $u, v \in \mathcal{F}_{A,b}$, then $\mathcal{M}_{\mathcal{W},\rho}$ is as well irreducible, aperiodic and its stationary distribution is ρ (Diaconis and Saloff-Coste, 1998, Lemma 1.1). Moreover, running a Metropolis-Hastings walk in practice is simple: Suppose the random walk is at node $u \in \mathcal{F}_{A,b}$, then we sample $v \in \mathcal{F}_{A,b}$ according to $\mathcal{W}(u, \cdot)$ and compute

$$p_{\mathcal{W}}(u, v) := \min \left\{ 1, \frac{\mathcal{W}(v, u)}{\mathcal{W}(u, v)} \cdot \prod_{i=1}^d \frac{u_i!}{v_i!} \right\}$$

where we used ρ from Equation 2.1. In a second step, we walk to v with probability p and we stay at u with probability $1 - p$ (see Algorithm 2 in Section A). Clearly, if $\mathcal{W}$ is symmetric, $p_{\mathcal{W}}(u, v)$ simplifies further. In order to construct a random walk $\mathcal{W}$ on $\mathcal{F}_{A,b}$ feasible for Metropolis-Hastings, we not only need irreducibility, i.e. all elements in the fiber can be reached, but also control over the transition probabilities to compute $p_{\mathcal{W}}(u, v)$. One way to construct a random walk is by using tools of commutative algebra as proposed in the seminal paper Diaconis and Sturmfels (1998), where one has to compute a *Markov basis* $\mathcal{M} \subset \mathbb{Z}^d$ for A , which is a set of vectors whose corresponding binomials in the polynomial ring constitute a generator set of the toric ideal of A . Geometrically, a Markov basis allows to connect any two elements in any fiber of a given matrix A . We denote the corresponding transition matrix with $\mathcal{W}_{\text{MB}}$. The obtained graph on the fiber is called *fiber graph*. Typically, scenarios with finite fibers are considered, where A can be chosen to have non-negative entries only.

3. FIBERS AS SAT-INSTANCES

In this section, we present our approach to represent $\mathcal{F}_{A,b}$ for given $A \in \mathbb{N}^{k \times d}$ and $b \in \mathbb{Z}^k$ as a SAT-instance. By definition, $\mathcal{F}_{A,b}$ is the set of all solutions $u \in \mathbb{N}^d$ to the k many

equations $A_i u = b_i$ and hence it can be seen as an instance of a constraint satisfaction problem over $\mathbb{Z}$ with positive solutions only. SAT-solvers are designed to solve constraint satisfaction problems over boolean algebras and hence their solution space is in $\{0, 1\}^q$ for some $q \in \mathbb{N}$. Equivalently, the solution space can be $[2^l - 1]^d$ for some $l \in \mathbb{N}$ by using l bits to represent each entry. Thus, if we would only consider solutions u whose entries have at most l -many bits, i.e. $u \in [2^l - 1]^d$, then generating elements from $\mathcal{F}_{A,b} \cap ([2^l - 1]^d)$ can be reduced to the satisfiability problem of boolean formulas in conjunctive normal form (CNF). The equations $Au = b$ can be reformulated as a series of additions, multiplications and equality checks. These can each be implemented by a boolean circuit C of size $\text{poly}(k, d, l)$ on variables $u_1, \dots, u_d$. Using the classic Tseitin encoding, C can be transformed to a CNF F which is satisfiable if and only if C is satisfiable. Moreover, $C(u_1, \dots, u_d) = 1$ if and only if there exists $y_1, \dots, y_m$ such that $F(u_1, \dots, u_d, y_1, \dots, y_m) = 1$. Particularly, $m \in O(|C|)$ and for every solution u to C the corresponding $y_1, \dots, y_m$ are unique. Thus, these transformations yield a bijection between $\mathcal{F}_{A,b} \cap [2^l - 1]^d$ and the satisfying solutions to F . It is easy to see that for $l = \lceil \log_2(\max_{i,j, A_{i,j} > 0} \frac{b_i}{A_{i,j}}) \rceil$ we have $\mathcal{F}_{A,b} = \mathcal{F}_{A,b} \cap ([2^l - 1]^d)$. For independence tests, A is a matrix with entries in $\{0, 1\}$ and thus we need at most as many bits as the largest entry of b . In other words, we can embed $\mathcal{F}_{A,b} \cap ([2^l - 1]^d)$ into the hypercube $\{0, 1\}^{l \cdot d + |C|}$. This allows us to use SAT-based techniques to generate fiber elements. Clearly, for small problem instances, the fiber can be enumerated and the p -value can be computed exactly.

Moving from generation of solutions to *uniform* sampling of solutions is more involved. Ideally, we would like to know the transition probabilities $\mathcal{W}_{\text{SAT}}(u, v)$ for two solutions explicitly, but this remains computationally intractable. Thus, we resort to SAT-samplers that are designed to sample uniformly or almost uniformly. However, it is not easy to obtain uniform sampling by just randomizing variable order or branching decisions of a CDCL solver. Also adding heuristics, like more preprocessing and clause learning, prohibits uniform sampling even more. Nevertheless, there exist SAT-solvers that can generate uniform or almost uniform random solutions to a given CNF. One is UniGen3 Soos et al. (2020), which uses universal hashing to randomly select areas of the solution space and generates one solution in this area. Another is CMSGen Golia et al. (2021), a Cryptominisat5 based SAT-sampler, which is built to fool the statistical test Barbarik Pote and Meel (2022b,a) for uniform sampling. We found CMSGen to be faster than UniGen3 while still being somehow uniform. Let $p(u)$ be the probability that a SAT-sampler draws the solution u from $\mathcal{F}_{A,b}$, then almost uniform samplers like UniGen3 satisfy $\frac{1-\epsilon}{|\mathcal{F}_{A,b}|} \leq p(u) \leq \frac{1+\epsilon}{|\mathcal{F}_{A,b}|}$ for a controllable hyperparameter $\epsilon > 0$. Therefore, the transition probabilities are all positive giving a complete graph on $\mathcal{F}_{A,b}$. On the other hand, CMSGen Golia et al. (2021) is designed to be not too far from a uniform sampler, i.e. the statistical test Barbarik did not conclude that the total variance distance to the uniform is more than η , that is, $\sum_{u \in \mathcal{F}_{A,b}} |p(u) - \frac{1}{|\mathcal{F}_{A,b}|}| \leq \eta$ is satisfied. Particularly, UniGen3 allows to control per-solution bounds while CMSGen only controls the total distance on the fiber. Golia et al. (2021) report that the Barbarik test could not distinguish between Unigen3 and CMSGen with $\eta = 1.8$. All SAT-samplers discussed in our work generate solutions that are statistically independent of previous sampled solutions.

4. DESIGNING SAT-BASED FIBER RANDOM WALKS

Integrating SAT-methods into the Metropolis-Hastings framework requires a careful design when convergence to a target distribution is desired. Particularly, this requires to control the individual transition probabilities $\mathcal{W}_{\text{SAT}}(u, v)$ between two fiber elements $u, v \in \mathcal{F}_{A,b}$. In general, these transition probabilities are unknown for SAT-samplers. Thus, we use SAT-samplers designed to approximate the uniform distribution. Naively, one could solely use a SAT-sampler to sample from the fiber $\mathcal{F}_{A,b}$ directly. In our experiments, however, we found that almost uniform sampling computationally is too expensive already for small problem instances. Although not-too-far-from uniform samplers like CMSGen allow to process large problem instances, their structural bias hinders a reliable p -value approximation already in small problem instances as shown in Figure 2. Consequently, to reduce structural bias but keep the good mixing properties, we can only use steps from the SAT-samplers selectively. To use them effectively in combination with existing methods, we use them to ensure irreducibility of the underlying Markov chain. Although Markov-basis methods can control the target distribution freely, they can quickly become large as they need to connect all fibers of a given constraint matrix A . Particularly for small right-hand sides and in the presence of structural zeros, sampling applicable moves from a Markov basis can become unlikely, leading to slow exploration of the fiber. Thus, our first hybrid method is the $A_n(\mathcal{M})$ scheme where every n -th step is a SAT-step while all others are from a predefined set of moves $\mathcal{M}$. Particularly, it alternates between steps from SAT-sampler and $\mathcal{M}$. The second scheme called $P_{n,k}(\mathcal{M})$ manages k -many walks in parallel. It first samples n -many independent initial points from the fiber using the SAT-sampler and then k -many random walks using $\mathcal{M}$ are executed. Note that we do not run these runs *in parallel* but sequentially: Every n steps, one of k -many walks is selected uniformly at random and continued for n steps. Clearly, if the total number of steps T satisfies $T = n \cdot k$, then both hybrid methods perform k -many SAT-steps and $T - k$ many steps using $\mathcal{M}$ and the only difference is the order at which the steps are generated, which can effect speed of mixing. Note that it is important to keep the ratio between SAT-steps and $\mathcal{M}$ -steps low to mitigate structural bias coming from the SAT-sampler.

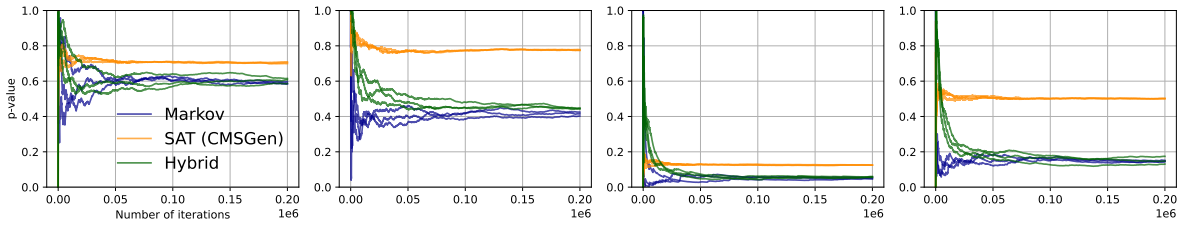


FIGURE 2. Convergence to p -values for independence testing in sparse 10×10 contingency tables, where SAT-only methods show strong structural biases prohibiting direct use for approximating p -values. In a hybrid combination with Markov basis, chains converge to correct p -value.

5. BENCHMARK STUDY

To measure effectiveness and efficiency of a sampling method, counting the number of samples generated by a sampler as done in prior work is not insightful when it comes to approximating p -values. For instance, a rapidly mixing Markov chain needs far fewer samples than a slowly mixing one. Instead, we empirically measure the convergence speed of the generated sequence $\frac{1}{j} \sum_{i=1}^j f(u_i)$ to the conditional p -value $\mathbb{E}_\rho[f]$ of Fishers test. Convergence speed depends on many factors, from dimensionality, sample size, to the actual initial observation. Thus we measure convergence for different constraint matrices, sample sizes, and initial observations. Different than existing work, we do not focus on the number of samples generated but on the number of steps required to reach a certain accuracy in terms of approximating the conditional p -value which is more meaningful in practice. Our evaluation scheme is shown in Algorithm 1 in detail. For a given constraint matrix A , we first construct initial samples $\{\hat{u}^1, \dots, \hat{u}^T\}$, where $T = 100$ in our experiments. Then we run Fishers test for each $\hat{u}^i$ using the Metropolis-Hastings walk for a predefined number of steps and measure the speed of their convergence to their limit. At the back end, Metropolis-Hastings uses a vanilla Markov basis sampler and the proposed SAT-sampler. In our experiments, the following model classes are considered:

- $\mathbf{I}_{d_1 \times \dots \times d_k}$: The $d_1 \times d_2 \times \dots \times d_k$ -independence model.
- $\mathbf{QI}_{d_1 \times \dots \times d_k}(S)$: $\mathbf{I}_{d_1 \times \dots \times d_k}$ with structural zeros $S \subseteq [d_1] \times \dots \times [d_k]$.
- $\mathbf{N3F}_d$: The no-3-factor interaction model on $d \times d \times d$ tables.

We explain these models in detail in Section 5.1. Note that the matrix of the sufficient statistics have their entries all in $\{0, 1\}$. In the reminder of this section, we state implementation details of Algorithm 1 for the different models used. We highlight that currently, we report the number of steps used by the sampler only, not the processing time required as this depends heavily on the system running the evaluation. Clearly, one benefit of samplers running with Markov bases are the low system requirements compared with SAT-samplers. Also, the samples/second of SAT-samplers depend on the problem size (number of boolean variables, number of clauses) and structure. On the other hand, computing a Markov basis beforehand is expensive and requires very high system requirements. Thus, a comparison in terms of processing time between Markov basis samplers and SAT samplers cannot be made fair. For $\mathbf{QI}_{5 \times 5}$, we were able to compute the exact conditional p -value by enumerating the whole fiber using our SAT-encoding. This allows us to measure the actual convergence speed to the exact p -value.

To be agnostic to the model family and structural zeros, we implemented a generic gradient-descent based MLE computation for log-linear models with constraint matrix $A \in \mathbb{N}^{k \times d}$ and structural zeros $S \subseteq [d]$ and initial observation $u^{\text{obs}} \in \mathbb{N}^d$. For a parameter $\theta \in \mathbb{R}^k$ we define probabilities on $\bar{S}$ by the softmax of the linear scores $\pi_\theta(i) \sim \exp(A^\top \theta)_i$ for $i \in \bar{S}$ and $\pi_\theta(i) = 0$ for $i \in S$. We want to compute $\hat{\theta}$ that maximizes the log-likelihood

$$\ell(\theta) = \sum_{i \in \bar{S}} u_i^{\text{obs}} \cdot \log \pi_\theta(i)$$

Algorithm 1: Evaluation scheme

Input : $A \in \mathbb{R}^{k \times d}$, number of runs T , sample size n , number of iterations N
Output: Sequences $p_1, \dots, p_T \in [0, 1]^N$
Generate initial values $\hat{u}^1, \dots, \hat{u}^T \in \mathbb{N}_0^d$ with $\|\hat{u}^i\|_1 = n$
for $i \in [T]$ **do**
 if *with structural zeros* **then**
 | Compute $S \subseteq \{j \in [d] : (\hat{u}^i)_j = 0\}$ such that MLE exists for $\mathcal{P}_{A,S}$
 else
 | $S = \emptyset$
 end
 Compute MLE $\tilde{\pi} \in \mathcal{P}_{A,S}$ for $\hat{u}^i$ under $\mathcal{P}_{A,S}$
 Define $X : \mathbb{R}^d \rightarrow \mathbb{R}_{\geq 0}$, $X(u) := \sum_{i=1}^d \frac{(\frac{u_i}{n} - \tilde{\pi}_i)^2}{\tilde{\pi}_i}$
 Run Metropolis-Hastings walk $(\hat{u}^i, X, N)$ and write result to $p_i \in \mathbb{R}^N$
end
return $p_1, \dots, p_T$

yielding $\hat{u}(\theta) := \|u^{\text{obs}}\|_1 \cdot \pi_\theta$ as MLE estimate. We optimize ℓ with L-BFGS on θ and monitor the margin discrepancy $\|A(u^{\text{obs}} - \hat{u}(\theta))\|_2$ during optimization and reduce the learning rate if the margin discrepancy does not decrease for a number of iterations.

5.1. Sufficient statistics matrices.

5.1.1. *Independence models.* Given a $d_1 \times \dots \times d_k$ contingency table u , the sufficient statistics matrices $A_{d_1, \dots, d_k}$ of $I_{d_1 \times \dots \times d_k}$ compute the $d_1 + d_2 + \dots + d_k$ many 1-marginals of u . In our experiments, we consider models on two- and three-way contingency tables, which are well studied in both, statistics and combinatorics. For instance, for the two-way independence model, A_{d_1, d_2} corresponds to the node-edge incidence matrix of the complete bipartite graph K_{d_1, d_2} and the elements of a Markov basis correspond to 4-cycles in K_{d_1, d_2} . The elements of this Markov basis are often called *basic moves* in the literature. In our experiments, we use the Markov basis consisting of basic moves for the Markov basis sampler.

If all 1-margins are strictly larger than zero, the MLE and a closed-form solutions exists, see for instance Yvonne M. M. Bishop (2007). For $I_{d_1 \times d_2}$, the 1-marginals are exactly the row sums $r_1, \dots, r_{d_1}$ and column sums $c_1, \dots, c_{d_2}$ and the MLE is $\tilde{\pi}_{i,j} := \frac{r_i \cdot c_j}{n \cdot n}$ where $n = \|u\|_1$ is the sample size (see for instance (Drton et al., 2009, Example 2.1.2)). Thus, when creating the initial tables, we ensure that the margins are positive. When creating the initial two-way tables $\hat{u}^i$, we make sure that we have varying distance to the expected frequencies $n \cdot \tilde{\pi}$. Particularly, we set $\hat{u}^i = \lambda \cdot u_{\text{indep}} + (1 - \lambda) \cdot u_{\text{dep}}$ where $\lambda \in [0, 1]$ and where u_{indep} and u_{dep} are randomly generated tables whose margins are independent and dependent respectively.

5.1.2. *Two-way quasi-independence model.* Here, the sufficient statistics is the same as for $I_{d_1 \times d_2}$, but additionally, a set $S \subset [d_1] \times [d_2]$ of structural zeros is given and a walk on

$$\mathcal{F}_{A,b} \cap \{u \in \mathbb{N}^d : u_s = 0 \ \forall s \in S\}$$

needs to be performed. Generally, the MLE does not exist for quasi-independence models making statistical analysis challenging Fienberg (1972). More about the uniqueness and existence of MLE in quasi-independence models can be found in Yvonne M. M. Bishop (2007). Recently, the authors of Coons and Sullivant (2021) gave an explicit formula for the MLE $\tilde{\theta}$ given an observation u^{obs} in cases the MLE exists and is rational (see (Coons and Sullivant, 2021, Theorem 5.4)). More specifically, their formula is valid for cases where the bipartite graph $K_{d_1, d_2} \setminus S$, that is the complete bipartite graph where the edges $(i, j) \in S$ have been removed, is *doubly chordal bipartite*. To generate the initial values in a way the MLE exists, we create an initial table u exactly as in Section 5.1.1 for the independence model. We then define the set of structural zeros initially to $S = \{(i, j) : \hat{u}_{i,j} = 0\}$ and checking whether $K_{d_1, d_2} \setminus S$ is doubly chordal bipartite. If not, we add a random edge to a cycle having less than two chords and extend S in that way until the graph is doubly chordal bipartite. Note that the basic moves are not guaranteed to connect all elements in $\mathcal{F}_{A,b} \cap \{u \in \mathbb{N}^d : \forall s \in S u_s = 0\}$. We refer to Aoki and (2005) and Hara et al. (2009) for a deeper analysis on when this is the case. Thus, we have to use the *Graver basis* (De Loera et al., 2012, Section 3) in general. Geometrically, the elements of the Graver basis corresponds to the set of all cycles in $K_{d_1, d_2} \setminus S$ and thus the Graver basis grows exponentially with the dimensions of the tables.

5.1.3. No-three way interaction model. Mathematical descriptions for Markov bases of no-three way interaction models only exist in special cases, like for $3 \times d \times d$ (see Aoki and Takemura (2003)). In our experiments, we used precomputed Markov bases from <https://markov-bases.de> for $3 \times 3 \times 3$ and $4 \times 4 \times 4$ consisting of 81 and 148.968 elements respectively. For a given sample size n , we sample n many indices from $[d] \times [d] \times [d]$ uniformly to obtain a random three-way table v^i for all $i \in [T]$. We then used the SAT-sampler to get an almost uniform element $\hat{u}^i$ from the fiber with margins $A \cdot v^i$.

5.2. Results. We observe that the proposed SAT-based random walks outperform Markov basis based random walks in many scenarios, especially in involved examples with small sample sizes or a large number of structural zeros. Selected results are shown in Figure 3a–3d. There, we show the number of steps required to be within 0.005 of the final limit. For smaller instances, like $\text{QI}_{5 \times 5}$, we compute convergence against the true p -value by enumerating the whole fiber using our SAT-encoding. Generally, we observe that our SAT-based methods converge faster than Markov basis based methods when tables are sparse, that is the sample size of the tables is small or structural zeros are present. This effect seems to be alleviated for larger sample sizes as the fibers become larger and typically better connected by local moves. In scenarios like $\text{I}_{k \times k}$, not much difference in convergence can be reported. Notably, for larger problem instances, $A_n(\mathcal{M})$ outperforms $P_{n,k}(\mathcal{M})$ as shown in Figure 3a, Figure 3b, and Figure 10. We observed that for quasi-independence models, there is no need to use a full Markov basis, like the Graver basis, to ensure convergence to the correct p -value. Surprisingly, we observe a faster convergence when using only the basic moves together with SAT-steps as shown in Figure 4. All results of our experiments are shown in the appendix in Section B.

6. CHANCES, RISKS AND LIMITATIONS

The SAT based perspective offers several chances for conditional inference on contingency tables. Most importantly, it enables methods that do not rely on Markov bases, which removes

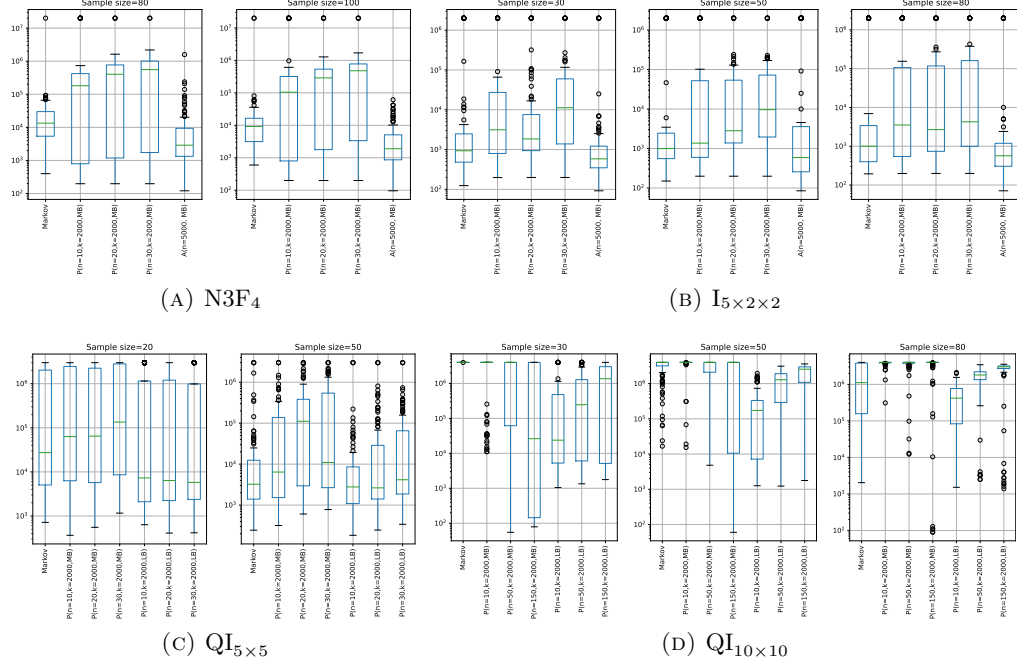


FIGURE 3. Selected results where SAT-based random walks outperform Markov bases based random walks.

a major computational barrier in models with many constraints and structural zeros. Modern SAT samplers can accelerate practical convergence by moving quickly across distant regions of the fiber, which reduces dependence between successive draws compared with purely local moves. When a lattice basis connects the fiber within slices, it can be used to generate simple proposals with low overhead, providing a practical alternative in settings where a full Markov basis is unknown or too costly. These advantages are most pronounced in involved examples with small sample sizes or a large number of structural zeros. In these cases, we think our method closes a particular performance gap, while other methods like Heat-bath based Markov chains Stanley and Windisch (2018) that dynamically adjust the lengths of the moves pay out their full potential in the limit of large sample sizes.

These gains come with risks that require careful mitigation. SAT samplers may introduce structural sampling errors which can distort the empirical distribution on the fiber and bias conditional p -values. Our hybrid-methods take care of these biases, but they also come with hyperparameters one may need to tune, like the number of walks or steps per walk. Finally, there are intrinsic limitations. The computation time per sample can be higher than for a single local move because sampling often involves repeated solver calls, for example with XOR constraints. However, the longer time can be a good investment, as subsequent samples from the SAT-sampler may come from fiber regions far apart where subsequent samples of Markov moves coincide in most of their entries.

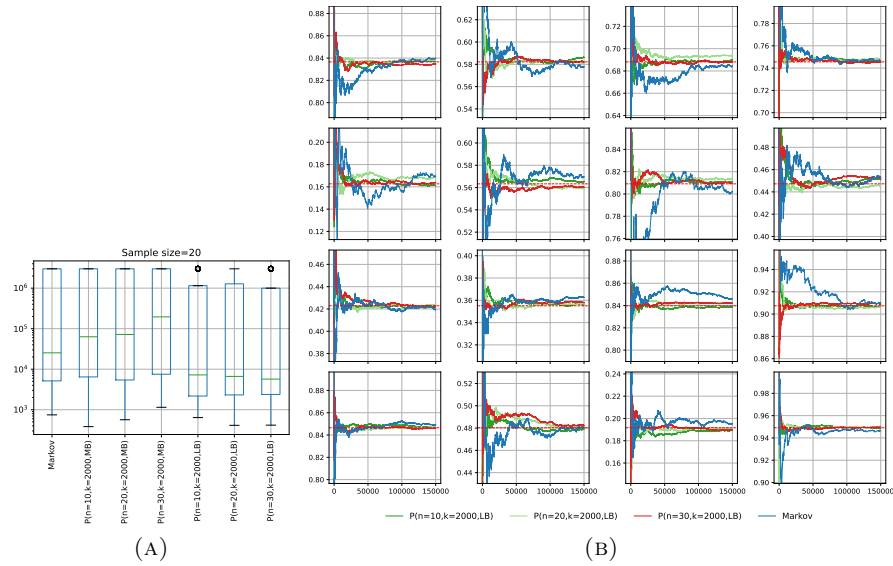


FIGURE 4. Number of steps to converge against true p -value (dotted) for $QI_{5 \times 5}$ using basic moves only.

For large design matrices and rich constraint sets, the Boolean encoding can grow rapidly, which increases memory usage and may render the SAT instance too large for practical solving. In such cases simpler lattice or Markov walks without SAT support or asymptotic approximations may be preferable for very large fibers, while exact extraction can dominate for very small fibers. Our implementation therefore includes fallbacks and tuning options, such as slice granularity and restart policies, to navigate these regimes, but it does not remove the fundamental scaling limits that arise from very high dimensional constraint systems.

Acknowledgements. The authors gratefully acknowledge the support received from the Hightech Agenda Bayern.

Data availability. The implemented algorithms are will be freely available under https://github.com/scharpfenecker/sat_fiber_walks on acceptance.

Conflict of interests. The authors provide no conflict of interest associated with the content of this article.

REFERENCES

- MeGASampler. <https://github.com/chaosite/MeGASampler>, 2025. Accessed: 2025-09-30.
- Satoshi Aoki and Akimichi Takemura and. Markov chain monte carlo exact tests for incomplete two-way contingency tables. *Journal of Statistical Computation and Simulation*, 75(10):787–812, 2005. doi: 10.1080/00949650410001690079. URL <https://doi.org/10.1080/00949650410001690079>.

- Satoshi Aoki and Akimichi Takemura. Minimal basis for a connected markov chain over $3 \times 3 \times k$ contingency tables with fixed two-dimensional marginals. *Australian & New Zealand Journal of Statistics*, 45:229 – 249, 06 2003. doi: 10.1111/1467-842X.00278.
- Miles Bakenhus and Sonja Petrović. Sampling lattice points in a polytope: A Bayesian biased algorithm with random updates. *Algebraic Statistics*, 15(1):61–83, 2024. doi: 10.2140/astat.2024.15.61. URL <https://doi.org/10.2140/astat.2024.15.61>.
- Armin Biere, Tobias Faller, Katalin Fazekas, Mathias Fleury, Nils Froleyks, and Florian Pollitt. CaDiCaL, Gimsatul, IsaSAT and Kissat entering the SAT Competition 2024. In Marijn Heule, Markus Iser, Matti Järvisalo, and Martin Suda, editors, *Proc. of SAT Competition 2024 – Solver, Benchmark and Proof Checker Descriptions*, volume B-2024-1 of *Department of Computer Science Report Series B*, pages 8–10, 2024.
- Jane Ivy Coons and Seth Sullivant. Quasi-independence models with rational maximum likelihood estimator. *Journal of Symbolic Computation*, 104:917–941, 2021. ISSN 0747-7171. doi: <https://doi.org/10.1016/j.jsc.2020.10.006>. URL <https://www.sciencedirect.com/science/article/pii/S0747717120301164>.
- Jesús A. De Loera, Raymond Hemmecke, and Matthias Köppe. *Algebraic and Geometric Ideas in the Theory of Discrete Optimization*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2012. doi: 10.1137/1.9781611972443. URL <https://epubs.siam.org/doi/abs/10.1137/1.9781611972443>.
- Persi Diaconis and Laurent Saloff-Coste. What Do We Know about the Metropolis Algorithm? *Journal of Computer and System Sciences*, 57(1):20–36, 1998. doi: 10.1006/jcss.1998.1576.
- Persi Diaconis and Bernd Sturmfels. Algebraic algorithms for sampling from conditional distributions. *The Annals of statistics*, 26(1):363–397, 1998.
- Adrian Dobra. Dynamic markov bases, 2011. URL <https://arxiv.org/abs/1103.4891>.
- Mathias Drton, Bernd Sturmfels, and Seth Sullivant. *Lectures on Algebraic Statistics*, volume 39 of *Oberwolfach Seminars*. 2009. doi: 10.1007/978-3-7643-8905-5.
- Rafael Dutra, Jonathan Bachrach, and Koushik Sen. Smtsamplers: Efficient stimulus generation from complex smt constraints. In *2018 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 1–8. IEEE, 2018a.
- Rafael Dutra, Kevin Laeufer, J. Bachrach, and Koushik Sen. Efficient sampling of sat solutions for testing. *2018 IEEE/ACM 40th International Conference on Software Engineering (ICSE)*, pages 549–559, 2018b. doi: 10.1145/3180155.3180248.
- Rafael Dutra, Jonathan Bachrach, and Koushik Sen. Guidedsampler: coverage-guided sampling of smt solutions. In *2019 Formal Methods in Computer Aided Design (FMCAD)*, pages 203–211. IEEE, 2019.
- Stephen E. Fienberg. The analysis of incomplete multi-way contingency tables. *Biometrics*, 28(1):177–202, 1972. ISSN 0006341X, 15410420. URL <http://www.jstor.org/stable/2528967>.
- Priyanka Golia, Mate Soos, Sourav Chakraborty, and Kuldeep S. Meel. Designing samplers is easy: The boon of testers. In *Proceedings of Formal Methods in Computer-Aided Design (FMCAD)*, 8 2021.
- Hisayuki Hara, Akimichi Takemura, and Ruriko Yoshida. Markov bases for two-way subtable sum problems. *Journal of Pure and Applied Algebra*, 213(8):1507–1521, 2009.

- ISSN 0022-4049. doi: <https://doi.org/10.1016/j.jpaa.2008.11.019>. URL <https://www.sciencedirect.com/science/article/pii/S002240490800251X>. Theoretical Effectivity and Practical Effectivity of Gröbner Bases.
- Hisayuki Hara, Satoshi Aoki, and Akimichi Takemura. Running markov chain without markov bases. 08 2011. doi: 10.1142/9789814383462_0005.
- W. K. Hastings. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57(1):97–109, 1970.
- Martin L. Hazelton and Masoud Karimi. When lattice bases are markov bases. *Statistics & Probability Letters*, 209:110106, 2024. ISSN 0167-7152. doi: <https://doi.org/10.1016/j.spl.2024.110106>. URL <https://www.sciencedirect.com/science/article/pii/S0167715224000750>.
- R. Heradio, David Fernández-Amorós, J. Galindo, David Benavides, and D. Batory. Uniform and scalable sampling of highly configurable systems. *Empirical Software Engineering*, 27, 2022. doi: 10.1007/s10664-021-10102-5.
- Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.
- Yash Pote and Kuldeep S. Meel. On scalable testing of samplers. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022a.
- Yash Pote and Kuldeep S Meel. On scalable testing of samplers. *Advances in Neural Information Processing Systems*, 35:28068–28079, 2022b.
- Fabio Rapallo. Markov bases and structural zeros. *Journal of Symbolic Computation*, 41(2):164–172, 2006. ISSN 0747-7171. doi: <https://doi.org/10.1016/j.jsc.2005.04.002>. URL <https://www.sciencedirect.com/science/article/pii/S0747717105001100>. Computational Algebraic Statistics.
- Subhajit Roy, Awanish Pandey, Brendan Dolan-Gavitt, and Yu Hu. Bug synthesis: Challenging bug-finding tools with deep faults. In *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, pages 224–234, 2018.
- Arijit Shaw and Kuldeep S Meel. Csb: A counting and sampling tool for bit-vectors. 2024.
- Mate Soos. The cryptominisat 5 set of solvers at sat competition 2016. *Proceedings of SAT Competition*, 28, 2016.
- Mate Soos, Stephan Gocht, and Kuldeep S Meel. Tinted, detached, and lazy cnf-xor solving and its applications to counting and sampling. In *International Conference on Computer Aided Verification*, pages 463–484. Springer, 2020.
- Niklas Sörensson. Minisat 2.2 and minisat++ 1.1. *A short description in SAT Race*, 2010, 2010.
- Caprice Stanley and Tobias Windisch. Heat-bath random walks with markov bases. *Advances in Applied Mathematics*, 92:122–143, 2018. ISSN 0196-8858. doi: <https://doi.org/10.1016/j.aam.2017.08.002>. URL <https://www.sciencedirect.com/science/article/pii/S0196885817301161>.

- Tobias Windisch. Rapid mixing and markov bases. *SIAM Journal on Discrete Mathematics*, 30(4):2130–2145, 2016. doi: 10.1137/15M1022045. URL <https://doi.org/10.1137/15M1022045>.
- Yang Xiang, Xiaowei Yang, Han Huang, Zhengxin Huang, and Miqing Li. Sampling configurations from software product lines via probability-aware diversification and sat solving. *Automated Software Engineering*, 29, 2022. doi: 10.1007/s10515-022-00348-8.
- Paul W. Holland Yvonne M. M. Bishop, Stephen E. Fienberg. *Maximum Likelihood Estimation For Incomplete Tables*. Springer New York, New York, NY, 2007. ISBN 978-0-387-72806-3. doi: 10.1007/978-0-387-72806-3_5.
- Ioannis Zachos, Mark Girolami, and Theodoros Damoulas. Generating origin-destination matrices in neural spatial interaction models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL <https://openreview.net/forum?id=eUcyIe1AzY>.
- Qingyang Zhang. Independence test for large sparse contingency tables based on distance correlation. *Statistics & Probability Letters*, 148:17–22, 2019. ISSN 0167-7152. doi: <https://doi.org/10.1016/j.spl.2018.12.010>. URL <https://www.sciencedirect.com/science/article/pii/S0167715218304048>.

APPENDIX A. METROPOLIS-HASTINGS ALGORITHM

Algorithm 2: Metropolis-Hastings for p -value approximation

Input : initial observation u^{obs} , number of steps N , random walk $\mathcal{W}$, extremeness measure X

Output: Sequences $p \in [0, 1]^N$

Set $u_0 = u^{\text{obs}}$ **for** $i \in [N]$ **do**

Sample $v \sim \mathcal{W}(u_{i-1}, \cdot)$

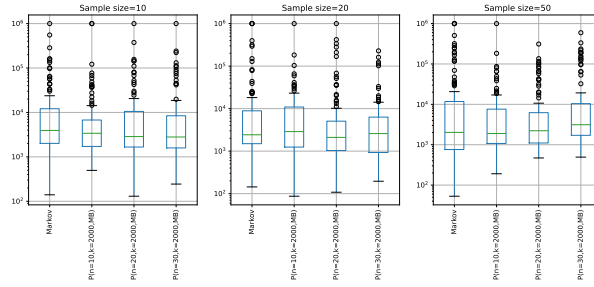
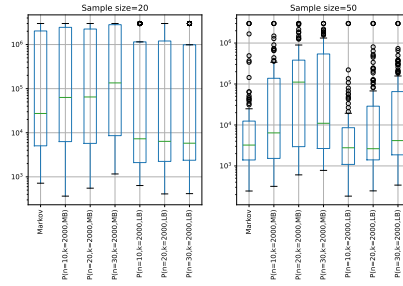
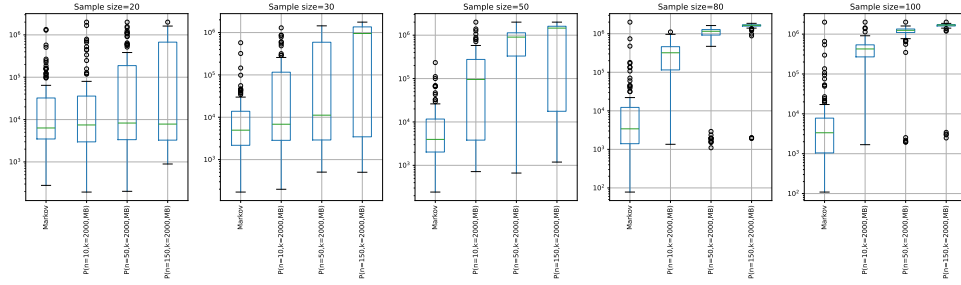
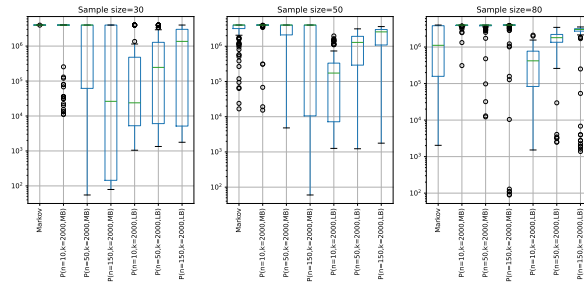
With probability $p_{\mathcal{W}}(u_{i-1}, v)$, set $u_i = v$, otherwise $u_i = u_{i-1}$

Compute $p_i = \frac{1}{i} \sum_{j=1}^i \mathbf{1}_{X(u_j) \geq X(u^{\text{obs}})}$

end

return p

APPENDIX B. FURTHER RESULTS


 FIGURE 5. Results for $I_{5 \times 5}$.

 FIGURE 6. Results for $QI_{5 \times 5}$.

 FIGURE 7. Results for $I_{10 \times 10}$.

 FIGURE 8. Results for $QI_{10 \times 10}$.

16 SAT-SAMPLING FOR STATISTICAL SIGNIFICANCE TESTING IN SPARSE CONTINGENCY TABLES

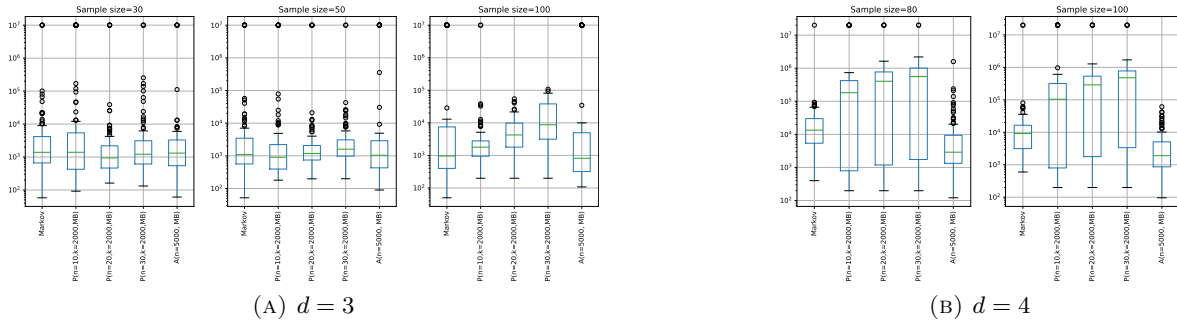


FIGURE 9. Results for $N3F_d$.

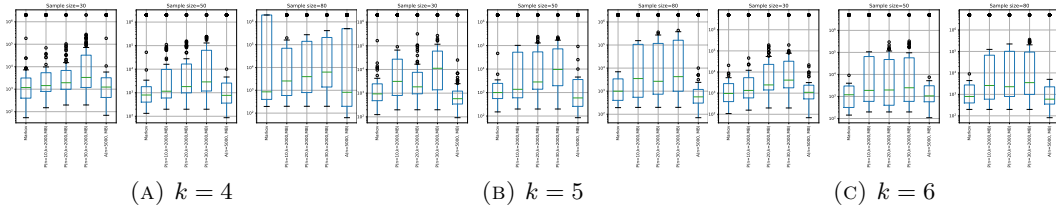


FIGURE 10. Results for $I_{k \times 2 \times 2}$.