A. MAP inference

Proposition 1 (MAP is divergence minimization)**.** *Let* $\Pr(\pi|\mathbf{C})$ *be a prior density over tables* $\pi \in \mathbb{R}^{m\times n}$*, satisfying the regularity properties* (3)*. Define* $Q(\pi) \triangleq -\log\Pr(\pi|\mathbf{C})$ *the negative log density. Then the posterior density* $\Pr(\pi|\mathbf{u},\mathbf{v},\mathbf{C})$ *has a unique maximum* π_* *which satisfies*

$$\pi_* = \operatorname*{argmin}_{\pi\in U(\mathbf{u},\mathbf{v})} \mathcal{D}_Q(\pi, \nabla Q^*(\mathbf{0})) \qquad (26)$$

with Q^* *the convex conjugate and* $\mathcal{D}_Q$ *the Bregman divergence with respect to* Q*.*

Proof. We first note that Q attains its global minimum at $\pi_* = \nabla Q^*(\mathbf{0}) \in \operatorname{int}(\operatorname{dom} Q)$, as the assumptions A1, A3 and A4 from (3) imply that $\nabla Q(\pi_*) = \mathbf{0}$ via the bijective relation $(\nabla Q)^{-1} = \nabla Q^*$, and Q is strictly convex on $\operatorname{int}(\operatorname{dom} Q)$, so the critical point is a global minimum.

The posterior density $\Pr(\pi|\mathbf{u},\mathbf{v},\mathbf{C})$ is the truncation of the prior $\Pr(\pi|\mathbf{C})$ to the polytope $U(\mathbf{u},\mathbf{v})$. Dessein (Dessein et al., 2018) in Section 3.2, Lemma 2, shows that the restriction of Q to $U(\mathbf{u},\mathbf{v})$ attains its global minimum uniquely at the Bregman projection of the unrestricted global minimum $\pi_* = \nabla Q^*(\mathbf{0})$ onto $U(\mathbf{u},\mathbf{v})$. The Lemma holds so long as $\nabla Q(\pi_*) = \mathbf{0}$ and $U(\mathbf{u},\mathbf{v}) \cap \operatorname{int}(\operatorname{dom} Q) \neq \emptyset$; the first is satisfied as noted above, and the second is satisfied by assumption A2 of (3). □

B. ϵ-MAP inference

Proposition 2 ($\epsilon-$MAP estimation)**.** *Let* $\Pr(\pi|\mathbf{C})$ *be a prior density over tables* $\pi \in \mathbb{R}^{m\times n}$*, satisfying the regularity properties A1, A2 and A3 from* (3)*. Define* $Q(\pi) \triangleq -\log\Pr(\pi|\mathbf{C})$ *the negative log density, and suppose there exists* $\pi_\epsilon \in \operatorname{int}(\operatorname{dom} Q)$ *such that* $\|\nabla Q(\pi_\epsilon)\|_2 < \epsilon$*. Let* π'_ϵ *be its Bregman projection,*

$$\pi'_\epsilon = \operatorname*{argmin}_{\pi\in U(\mathbf{u},\mathbf{v})} \mathcal{D}_Q(\pi, \pi_\epsilon). \qquad (27)$$

Then the posterior density $\Pr(\pi|\mathbf{u},\mathbf{v},\mathbf{C})$ *has a unique maximum* π_* *that satisfies*

$$Q(\pi'_\epsilon) - Q(\pi_*) < \sqrt{2}\epsilon. \qquad (28)$$

Proof. Let π'_ϵ be the Bregman projection of π_ϵ onto $U(\mathbf{u},\mathbf{v})$ with respect to Q. $U(\mathbf{u},\mathbf{v})$ is a closed, convex set, and assumption A2 of (3) implies that $U(\mathbf{u},\mathbf{v}) \cap \operatorname{int}(\operatorname{dom} Q) \neq \emptyset$, so the Bregman projection is well-defined. The Bregman projection is characterized by the relation

$$\langle \pi - \pi'_\epsilon, \nabla Q(\pi_\epsilon) - \nabla Q(\pi'_\epsilon)\rangle \leq 0, \qquad (29)$$

for all $\pi \in U(\mathbf{u},\mathbf{v}) \cap \operatorname{int}(\operatorname{dom} Q)$. From the definition of the Bregman divergence, we have that $\mathcal{D}_Q(\pi, \pi'_\epsilon) > 0$ for all $\pi \in \operatorname{int}(\operatorname{dom} Q)$, so

$$\begin{aligned} Q(\pi) - Q(\pi'_\epsilon) &> \langle \pi - \pi'_\epsilon, \nabla Q(\pi'_\epsilon)\rangle \\ &\geq \langle \pi - \pi'_\epsilon, \nabla Q(\pi_\epsilon)\rangle, \end{aligned}$$

with the second inequality deriving from (29). Q is strictly convex on $\operatorname{int}(\operatorname{dom} Q)$ and $U(\mathbf{u},\mathbf{v})$ is closed and convex, so Q has a unique minimum on $U(\mathbf{u},\mathbf{v}) \cap \operatorname{int}(\operatorname{dom} Q)$. Let π_* be this minimum. Inverting the previous inequality, we have

$$\begin{aligned} Q(\pi'_\epsilon) - Q(\pi_*) &< \langle \pi'_\epsilon - \pi_*, \nabla Q(\pi_\epsilon)\rangle \\ &\leq \|\pi'_\epsilon - \pi_*\|_2 \|\nabla Q(\pi_\epsilon)\|_2, \end{aligned}$$

by Cauchy-Schwarz. By assumption $\|\nabla Q(\pi_\epsilon)\|_2 < \epsilon$, while π'_ϵ and π_* both lie in the simplex $\Delta^{m\times n}$, meaning that $\|\pi'_\epsilon - \pi_*\|_2 \leq \sqrt{2}$. Combining these yields the bound (28). □

C. MAP inference with noisy observations

Proposition 3 (MAP with noise is a generalized projection)**.** *Let* $Q(\pi) \triangleq -\log\Pr(\pi|\mathbf{C})$*,* $\psi_{\mathbf{u}}(\pi) \triangleq -\log\Pr(\mathbf{u}|\pi\mathbf{1})$*, and* $\psi_{\mathbf{v}}(\pi) = -\log\Pr(\mathbf{v}|\pi^\intercal\mathbf{1})$ *be the negative log densities. Then the posterior density* $\Pr(\pi|\mathbf{u},\mathbf{v},\mathbf{C})$ *has a unique global maximum which satisfies*

$$\pi_* = \operatorname*{argmin}_{\pi\in\Delta^{m\times n}} \psi_{\mathbf{u}}(\pi) + \psi_{\mathbf{v}}(\pi) + \mathcal{D}_Q(\pi, \nabla Q^*(\mathbf{0})), \quad (30)$$

with Q^* *the convex conjugate of* Q *and* $\mathcal{D}_Q$ *the Bregman divergence with respect to* Q*.*

Proof. Define $f(\pi) = \psi_{\mathbf{u}^{(i)}}(\pi) + \psi_{\mathbf{v}^{(i)}}(\pi) + \mathcal{D}_Q(\pi, \nabla Q^*(\mathbf{0}))$, with domain $\operatorname{dom}\psi_{\mathbf{u}^{(i)}} \cap \operatorname{dom}\psi_{\mathbf{v}^{(i)}} \cap \operatorname{dom} Q$. Expanding $\mathcal{D}_Q$, we see that

$$\begin{aligned} &\mathcal{D}_Q(\pi, \nabla Q^*(\mathbf{0})) \\ &\quad= Q(\pi) - Q(\nabla Q^*(\mathbf{0})) - \langle \nabla Q(\nabla Q^*(\mathbf{0})), \pi - \nabla Q^*(\mathbf{0})\rangle \\ &\quad= Q(\pi) - Q(\nabla Q^*(\mathbf{0})), \end{aligned} \qquad (31)$$

because $\nabla Q^*(\mathbf{0}) \in \operatorname{int}(\operatorname{dom} Q)$ and Q is Legendre, so that ∇Q and ∇Q^* are inverses. So the restriction of f to $\Delta^{m\times n}$ differs from the log posterior by a constant $-\log\Pr(\mathbf{u},\mathbf{v}|\mathbf{C}) - Q(\nabla Q^*(\mathbf{0}))$.

By Assumption A2 of (3) and the assumption in Section 5.2, we have $\Delta^{m\times n} \cap \operatorname{int}(\operatorname{dom} f) \neq \emptyset$. Moreover, f is closed, strictly convex and coercive, while $\Delta^{m\times n}$ is closed and convex, so f has a unique minimum in $\Delta^{m\times n}$, and the same holds for the log posterior. □

D. Significance of empirical error differences

Each of the datasets in Section 6.2 consists of a number of tables whose values we attempt to infer, using the given models. Table 1 reports the median absolute error of these inferred tables, across all tables and table entries, and highlights the best performing method for each dataset. Since Table 1 only gives a single statistic for each set of inferred tables, there remains a question of significance of the differences between the algorithms.

As a first measure of this significance, we report here the percentage of tables for which each given method actually matches or outperforms the best performing method from Table 1, for each dataset. Specifically, for each table in the dataset, we compare the median absolute error within that table, for the given method, to the same for the best performing method, and report the percentage of cases in which the latter equals or exceeds the former.

Table 2 shows these percentages. They are very small ($<$ 0.1) in the overwhelming majority of cases, with several notable exceptions that approach or exceed 0.5 – of these, all but one fail the Wilcoxon test (decribed below), and these instances are bolded in the table.

As a second measure of significance, we compute p-values under the hypothesis that a given method actually matches or outperforms the best performing method from Table 1 on average. For each table in the dataset, we take the difference between the median absolute error within that table, for the given method, and the same for the best performing method. We test the hypothesis that the pseudomedian of these differences (across tables) is actually nonpositive, which would indicate that the differences are actually distributed significantly at or to the left of zero. This is the one-sided Wilcoxon signed rank test (Wilcoxon, 1945).

Table 3 shows the resulting $\log_{10} p$-values. These p-values are very small, with a handful of exceptions, indicating that in the overwhelming majority of cases the differences in accuracy between the different algorithms are significant. The exceptions are bolded in this table and in Table 1 in the main text, indicating cases where we cannot say definitively that the best-performing method in Table 1 outperformed the given method.

Table 2. Percentage of tables for which the median absolute error equals or is lower than that for the best performing method from Table 1. Bolded are instances that fail the Wilcoxon test (Table 3).

DATASET	DEATH	EDUCATION	VOTING	INCOME	INSURANCE
N. TABLES	51	110	68	51	51
Prior work					
NEIGHBORHOOD	0	0.01	0.31	0.02	0
GOODMAN	0	0	0	0	0
MULT. DIRICH. (MCMC)	0	0.01	–	0	0
MAP, exact					
ENTROPIC	0.04	0.6	0	0.04	0
TSALLIS ($q = 0.5$)	0	0.05	0.28	0.27	0.06
TSALLIS ($q = 2$)	0	0.05	0.04	0	0
NORMAL	–	0.16	0.07	–	–
DIRICHLET	0	0.05	0.09	0	**0.39**
MAP, multinomial					
ENTROPIC	0.12	–	0.01	0.18	0
TSALLIS ($q = 0.5$)	0	0.01	0.07	**0.51**	0.04
TSALLIS ($q = 2$)	0	0.02	0	0	0
NORMAL	0.02	0.09	0.21	**0.59**	0.18
DIRICHLET	0.12	0.01	0.01	0.18	**0.51**

Table 3. $\log_{10}$ p-value, Wilcoxon signed rank test (one-sided), differences between median absolute errors per-table, for given method vs. the best performing method from Table 1. Bolded are instances that fail the test ($p > 1e{-}3$).

DATASET	DEATH	EDUCATION	VOTING	INCOME	INSURANCE
N. TABLES	51	110	68	51	51
Prior work					
NEIGHBORHOOD	-9.59	-18.1	-3.02	-9.36	-9.59
GOODMAN	-9.59	-19.4	-12.4	-9.59	-9.59
MULT. DIRICH. (MCMC)	-9.59	-19.2	–	-9.59	-9.59
MAP, exact					
ENTROPIC	-9.26	-8.42	-12.4	-6.88	-9.59
TSALLIS ($q = 0.5$)	-9.59	-17.6	-3.39	-3.78	-7.48
TSALLIS ($q = 2$)	-9.59	-17.9	-11.6	-9.59	-9.59
NORMAL	–	-15.3	-11.5	–	–
DIRICHLET	-9.59	-17.3	-10.2	-9.59	**-1.54**
MAP, multinomial					
ENTROPIC	-7.91	–	-12.2	-6.88	-9.59
TSALLIS ($q = 0.5$)	-9.59	-17.6	-11.4	**-0.07**	-8.98
TSALLIS ($q = 2$)	-9.59	-18.1	-12.4	-9.59	-9.59
NORMAL	-9.56	-17.8	-7.77	**-0.21**	-6.06
DIRICHLET	-7.87	-19.0	-12.2	-6.92	**-0.27**