Table 1: Generalized Linear Models: Comparing the in-context learner with a Gaussian approximation, fitted via the forward KL-divergence, to the proposed flow matching method. Evaluation on 50 synthetic and 17 real-world datasets for seven different scenarios. If one method is by more than two standard errors better than the other, it is marked in **bold**. Overall, the ICL + Flow Matching method clearly outperforms the Gaussian approximation, fitted via the forward KL-divergence,: it yields significantly better results (according to the two-standard-error criterion) in 6 out of 7 scenarios on synthetic datasets and in all 7 scenarios on real-world datasets, across at least two of the three considered metrics (C2ST, MMD, or $\mathcal{W}_2$). In addition, the flow matching method consistently achieves lower or comparable standard errors, indicating more stable and reliable performance across datasets.

Scenario	**Model**	**Synthetic Evaluation**			**Real-World Evaluation**		
		C2ST (↓)	MMD (↓)	$\mathcal{W}_2$ (↓)	C2ST (↓)	MMD (↓)	$\mathcal{W}_2$ (↓)
Scenario 1	ICL + Gaussian	**0.845** (± 0.213)	**1.601** (± 1.213)	2.024 (± 0.874)	0.980 (± 0.007)	1.715 (± 0.295)	1.976 (± 0.238)
	ICL + Flow Matching	**0.765** (± 0.123)	**0.767** (± 0.727)	**0.585** (± 0.301)	**0.614** (± 0.074)	**0.175** (± 0.219)	**0.310** (± 0.138)
Scenario 2	ICL + Gaussian	**0.941** (± 0.056)	**1.000** (± 0.953)	1.943 (± 0.657)	0.969 (± 0.013)	1.490 (± 0.310)	2.068 (± 0.259)
	ICL + Flow Matching	**0.839** (± 0.072)	**0.707** (± 0.658)	**1.111** (± 0.300)	**0.768** (± 0.033)	**0.143** (± 0.089)	**0.411** (± 0.094)
Scenario 3	ICL + Gaussian	0.907 (± 0.138)	1.779 (± 1.363)	4.713 (± 1.560)	0.985 (± 0.006)	1.526 (± 0.198)	4.144 (± 0.438)
	ICL + Flow Matching	**0.611** (± 0.070)	**0.089** (± 0.114)	**0.423** (± 0.348)	**0.576** (± 0.027)	**0.037** (± 0.026)	**0.257** (± 0.044)
Scenario 4	ICL + Gaussian	0.989 (± 0.011)	3.544 (± 0.343)	23.035 (± 6.549)	0.990 (± 0.003)	3.858 (± 0.061)	13.601 (± 0.427)
	ICL + Flow Matching	**0.753** (± 0.049)	**0.171** (± 0.153)	**0.631** (± 0.294)	**0.762** (± 0.015)	**0.105** (± 0.046)	**0.597** (± 0.104)
Scenario 5	ICL + Gaussian	0.962 (± 0.037)	1.444 (± 1.640)	3.299 (± 1.614)	0.991 (± 0.005)	1.666 (± 0.387)	2.963 (± 0.239)
	ICL + Flow Matching	**0.621** (± 0.063)	**0.067** (± 0.080)	**0.299** (± 0.195)	**0.610** (± 0.045)	**0.046** (± 0.020)	**0.242** (± 0.038)
Scenario 6	ICL + Gaussian	0.909 (± 0.048)	1.020 (± 0.505)	1.515 (± 0.358)	0.939 (± 0.047)	1.799 (± 0.751)	1.904 (± 0.541)
	ICL + Flow Matching	**0.532** (± 0.019)	**0.016** (± 0.008)	**0.590** (± 0.066)	**0.556** (± 0.017)	**0.035** (± 0.015)	**0.504** (± 0.038)
Scenario 7	ICL + Gaussian	0.970 (± 0.030)	2.169 (± 1.473)	1.707 (± 0.480)	0.993 (± 0.006)	2.390 (± 0.414)	1.362 (± 0.152)
	ICL + Flow Matching	**0.700** (± 0.116)	**0.317** (± 0.355)	**0.400** (± 0.286)	**0.773** (± 0.048)	**0.294** (± 0.457)	**0.559** (± 0.256)

Table 2: Factor Analysis: Comparing the in-context learner with a Gaussian approximation, fitted via the forward KL-divergence, to the proposed flow matching method. Evaluation on 50 synthetic and 17 real-world datasets for seven different scenarios. If one method is by more than two standard errors better than the other, it is marked in **bold**. The flow matching approach shows favorable performance in the majority of cases. Specifically, it achieves statistically significant improvements in all 6 scenarios on synthetic data and in 5 out of 6 scenarios on real-world data. Notably, it often reduces discrepancy measures such as MMD and Wasserstein-2 distance by a large margin. In addition, the variability of the flow matching estimates is generally lower, leading to more reliable and consistent results across different datasets. In scenario 4, the Gaussian in-context learner learned a singular covariance matrix.

Scenario	**Model**	**Synthetic Evaluation**			**Real-World Evaluation**		
		C2ST (↓)	MMD (↓)	$\mathcal{W}_2$ (↓)	C2ST (↓)	MMD (↓)	$\mathcal{W}_2$ (↓)
Scenario 1	ICL + Gaussian	0.974 (± 0.028)	1.838 (± 0.778)	1.450 (± 0.607)	**0.589** (± 0.015)	**0.080** (± 0.010)	0.459 (± 0.017)
	ICL + Flow Matching	**0.552** (± 0.028)	**0.034** (± 0.034)	**0.289** (± 0.083)	**0.606** (± 0.038)	**0.068** (± 0.069)	**0.265** (± 0.078)
Scenario 2	ICL + Gaussian	0.835 (± 0.040)	0.813 (± 0.276)	1.250 (± 0.316)	0.889 (± 0.027)	0.778 (± 0.109)	1.074 (± 0.073)
	ICL + Flow Matching	**0.542** (± 0.006)	**0.017** (± 0.006)	**0.244** (± 0.033)	**0.622** (± 0.032)	**0.098** (± 0.039)	**0.287** (± 0.046)
Scenario 3	ICL + Gaussian	0.826 (± 0.035)	0.826 (± 0.226)	1.210 (± 0.239)	0.942 (± 0.008)	1.466 (± 0.078)	1.317 (± 0.038)
	ICL + Flow Matching	**0.537** (± 0.023)	**0.024** (± 0.021)	**0.259** (± 0.088)	**0.609** (± 0.019)	**0.124** (± 0.037)	**0.179** (± 0.018)
Scenario 4	ICL + Gaussian	0.870 (± 0.043)	0.706 (± 0.218)	1.635 (± 0.297)	0.999 (± 0.001)	2.025 (± 0.017)	2.013 (± 0.019)
	ICL + Flow Matching	**0.684** (± 0.060)	**0.198** (± 0.141)	**0.918** (± 0.246)	**0.988** (± 0.003)	**1.764** (± 0.026)	**1.248** (± 0.008)
Scenario 5	ICL + Gaussian	0.838 (± 0.029)	0.831 (± 0.219)	1.248 (± 0.249)	0.944 (± 0.009)	1.477 (± 0.073)	1.316 (± 0.031)
	ICL + Flow Matching	**0.535** (± 0.016)	**0.021** (± 0.011)	**0.279** (± 0.060)	**0.886** (± 0.017)	**1.207** (± 0.101)	**1.002** (± 0.042)
Scenario 6	ICL + Gaussian	0.837 (± 0.030)	0.831 (± 0.219)	1.248 (± 0.249)	0.944 (± 0.008)	1.477 (± 0.073)	1.316 (± 0.031)
	ICL + Flow Matching	**0.543** (± 0.021)	**0.023** (± 0.015)	**0.345** (± 0.173)	**0.666** (± 0.020)	**0.200** (± 0.034)	**0.224** (± 0.014)

Table 3: Gaussian Mixture Models: Comparing the in-context learner with a Gaussian approximation, fitted via the forward KL-divergence, to the proposed flow matching method. Evaluation on 50 synthetic and 17 real-world datasets for seven different scenarios. If one method is by more than two standard errors better than the other, it is marked in **bold**. While the differences are less clear-cut than in the previous models, ICL + Flow Matching demonstrates favorable performance in several scenarios, particularly for the Wasserstein-2 distance and MMD. Notably, its advantage is most visible in lower-dimensional settings (Scenario 1 and 2), where it consistently improves upon the Gaussian approximation, fitted via the forward KL-divergence, across most metrics. However, as the dimensionality increases, the performance gap tends to narrow, and in some cases, the inherent variability of the datasets, especially for the Gaussian approximation, fitted via the forward KL-divergence,, makes it difficult to conclusively determine a clear winner. Nonetheless, the flow matching approach often achieves smaller standard errors and lower discrepancy measures, underlining its potential for more stable modeling.

Scenario	**Model**	**Synthetic Evaluation**			**Real-World Evaluation**		
		C2ST (↓)	MMD (↓)	$\mathcal{W}_2$ (↓)	C2ST (↓)	MMD (↓)	$\mathcal{W}_2$ (↓)
Scenario 1	ICL + Gaussian	**0.926** (± 0.029)	**0.555** (± 0.452)	**2.586** (± 0.560)	**0.957** (± 0.034)	**0.765** (± 0.958)	**3.717** (± 1.709)
	ICL + Flow Matching	**0.760** (± 0.092)	**0.303** (± 0.548)	**2.095** (± 1.692)	**0.847** (± 0.082)	**0.486** (± 0.623)	**4.054** (± 2.782)
Scenario 2	ICL + Gaussian	0.985 (± 0.010)	0.761 (± 0.227)	5.022 (± 0.945)	**0.999** (± 0.001)	0.801 (± 0.256)	7.525 (± 1.513)
	ICL + Flow Matching	**0.812** (± 0.061)	**0.159** (± 0.154)	**2.314** (± 0.926)	**0.937** (± 0.041)	**0.282** (± 0.131)	**3.947** (± 1.055)
Scenario 3	ICL + Gaussian	**0.998** (± 0.002)	**0.829** (± 0.241)	**11.536** (± 2.365)	**1.000** (± 0.000)	**1.500** (± 0.251)	**26.242** (± 4.171)
	ICL + Flow Matching	**1.000** (± 0.000)	**0.582** (± 0.280)	**8.708** (± 4.945)	**1.000** (± 0.000)	**1.869** (± 0.342)	**33.230** (± 8.095)
Scenario 4	ICL + Gaussian	**0.998** (± 0.001)	**6.314** (± 0.449)	**13.404** (± 0.609)	**0.997** (± 0.001)	**2.770** (± 1.201)	22.596 (± 5.717)
	ICL + Flow Matching	1.000 (± 0.000)	**2.451** (± 0.868)	**8.333** (± 4.202)	1.000 (± 0.000)	**2.518** (± 0.694)	**11.938** (± 2.956)