Table 1: Generalized Linear Models: Ablation with respect to the dimensionality of the problem on 50 synthetic and 17 real-world datasets for scenarios 2 and 3. All results within two standard errors of the best average result for each scenario are marked in **bold**. Due to the limitations of the number of features in the real-world data, we can only use 5 datasets for 20 and one dataset for 50 dimensions. Overall, we find that the advantages of the in-context learning approach to deteriorate for higher dimensionalities, with the variational inference methods using a Gaussian approximation performing well for 20 dimensions. This finding is line with work by (Mittal et al., 2025a;b). For 50 dimensions we find that in many cases the used metrics do not allow to significantly discriminate the performance of the different approaches.

Scenario	Dim.	Model	Synthetic Evaluation			Real-World Evaluation		
			C2ST (↓)	MMD (↓)	$\mathcal{W}_2$ (↓)	C2ST (↓)	MMD (↓)	$\mathcal{W}_2$ (↓)
Scenario 2	5	Laplace Approximation	1.000 (± 0.000)	4.853 (± 2.333)	5.770 (± 5.946)	1.000 (± 0.000)	2.572 (± 0.206)	0.809 (± 0.149)
		VI: DiagonalNormal	0.957 (± 0.091)	3.906 (± 2.679)	5.628 (± 6.092)	0.892 (± 0.044)	0.847 (± 0.389)	**0.530** (± 0.175)
		VI: MultivariateNormal	0.910 (± 0.131)	3.407 (± 2.781)	5.584 (± 6.104)	0.820 (± 0.031)	0.243 (± 0.148)	**0.408** (± 0.118)
		VI: Structured Normal	0.908 (± 0.119)	3.139 (± 2.763)	5.480 (± 6.164)	0.824 (± 0.023)	0.215 (± 0.110)	**0.392** (± 0.109)
		VI: IAF	0.968 (± 0.063)	4.416 (± 2.473)	7.474 (± 6.235)	0.888 (± 0.067)	0.921 (± 0.860)	0.942 (± 0.733)
		ICL (ours)	**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 2	20	Laplace Approximation	1.000 (± 0.000)	2.314 (± 0.237)	3.069 (± 1.168)	1.000 (± 0.000)	2.222 (± 0.018)	2.847 (± 0.305)
		VI: DiagonalNormal	0.904 (± 0.168)	1.292 (± 0.937)	**2.863** (± 0.919)	0.990 (± 0.009)	1.277 (± 0.452)	2.483 (± 0.318)
		VI: MultivariateNormal	0.851 (± 0.134)	**0.492** (± 0.547)	**2.694** (± 0.916)	0.843 (± 0.069)	0.243 (± 0.170)	**2.166** (± 0.266)
		VI: Structured Normal	**0.697** (± 0.065)	**0.070** (± 0.099)	**2.497** (± 0.993)	**0.655** (± 0.031)	**0.029** (± 0.025)	**2.191** (± 0.271)
		VI: IAF	0.916 (± 0.110)	1.062 (± 1.076)	4.191 (± 0.623)	0.952 (± 0.025)	0.515 (± 0.242)	3.331 (± 0.371)
		ICL (ours)	0.955 (± 0.057)	1.131 (± 1.035)	4.945 (± 0.836)	0.968 (± 0.020)	0.724 (± 0.278)	4.356 (± 0.302)
Scenario 2	50	Laplace Approximation	**1.000** (± 0.000)	2.437 (± 0.271)	**5.728** (± 1.358)	1.000 (± nan)	2.350 (± nan)	5.620 (± nan)
		VI: DiagonalNormal	**0.853** (± 0.182)	0.787 (± 0.687)	**6.224** (± 1.225)	0.996 (± nan)	1.080 (± nan)	5.426 (± nan)
		VI: MultivariateNormal	**0.878** (± 0.150)	**0.688** (± 0.620)	**6.206** (± 1.244)	0.994 (± nan)	0.791 (± nan)	5.305 (± nan)
		VI: Structured Normal	**0.865** (± 0.081)	**0.186** (± 0.169)	5.874 (± 1.233)	0.819 (± nan)	0.093 (± nan)	5.660 (± nan)
		VI: IAF	**0.909** (± 0.130)	0.649 (± 0.650)	7.465 (± 0.335)	0.985 (± nan)	0.426 (± nan)	6.426 (± nan)
		ICL (ours)	**0.972** (± 0.039)	0.741 (± 0.713)	8.313 (± 0.608)	0.971 (± nan)	0.405 (± nan)	7.718 (± nan)
Scenario 3	5	Laplace Approximation	1.000 (± 0.000)	2.203 (± 0.997)	1.170 (± 0.949)	1.000 (± 0.000)	1.841 (± 0.185)	0.729 (± 0.175)
		VI: DiagonalNormal	0.866 (± 0.101)	1.069 (± 1.150)	0.846 (± 0.747)	0.797 (± 0.083)	0.526 (± 0.361)	0.480 (± 0.207)
		VI: MultivariateNormal	**0.656** (± 0.131)	**0.445** (± 1.061)	**0.660** (± 0.737)	**0.560** (± 0.035)	**0.032** (± 0.028)	**0.249** (± 0.069)
		VI: Structured Normal	**0.653** (± 0.125)	**0.421** (± 0.993)	**0.659** (± 0.736)	**0.552** (± 0.028)	**0.027** (± 0.015)	**0.239** (± 0.055)
		VI: IAF	0.751 (± 0.148)	0.939 (± 1.349)	0.964 (± 0.924)	0.673 (± 0.141)	0.399 (± 0.543)	0.563 (± 0.433)
		ICL (ours)	**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 3	20	Laplace Approximation	1.000 (± 0.000)	2.726 (± 1.116)	4.127 (± 1.927)	1.000 (± 0.000)	2.234 (± 0.092)	3.589 (± 0.519)
		VI: DiagonalNormal	**0.912** (± 0.134)	**1.704** (± 1.467)	**3.933** (± 1.574)	0.983 (± 0.014)	1.298 (± 0.443)	3.147 (± 0.557)
		VI: MultivariateNormal	**0.863** (± 0.113)	**0.937** (± 1.174)	**3.754** (± 1.650)	**0.796** (± 0.099)	**0.268** (± 0.226)	**2.645** (± 0.466)
		VI: Structured Normal	**0.768** (± 0.109)	**0.302** (± 0.518)	**3.151** (± 1.663)	**0.722** (± 0.073)	**0.131** (± 0.141)	**2.579** (± 0.399)
		VI: IAF	**0.908** (± 0.133)	**1.657** (± 1.476)	5.543 (± 1.120)	0.936 (± 0.041)	0.548 (± 0.341)	3.678 (± 0.670)
		ICL (ours)	**0.902** (± 0.076)	**1.053** (± 0.782)	6.206 (± 0.783)	0.932 (± 0.019)	0.635 (± 0.183)	5.281 (± 0.317)
Scenario 3	50	Laplace Approximation	**1.000** (± 0.000)	2.700 (± 0.789)	**8.841** (± 1.691)	1.000 (± nan)	2.348 (± nan)	7.049 (± nan)
		VI: DiagonalNormal	**0.870** (± 0.127)	**1.154** (± 1.321)	**9.180** (± 1.513)	0.997 (± nan)	1.393 (± nan)	6.791 (± nan)
		VI: MultivariateNormal	**0.896** (± 0.101)	**1.027** (± 1.157)	**9.175** (± 1.555)	0.998 (± nan)	1.092 (± nan)	6.667 (± nan)
		VI: Structured Normal	**0.873** (± 0.112)	**0.539** (± 0.667)	**9.118** (± 1.538)	0.958 (± nan)	0.420 (± nan)	6.665 (± nan)
		VI: IAF	**0.869** (± 0.124)	**0.751** (± 0.939)	**9.917** (± 0.870)	0.971 (± nan)	0.417 (± nan)	7.411 (± nan)
		ICL (ours)	**0.931** (± 0.062)	**0.784** (± 0.884)	10.063 (± 0.930)	0.965 (± nan)	0.347 (± nan)	8.482 (± nan)
Scenario 5	5	Laplace Approximation	1.000 (± 0.000)	2.060 (± 0.472)	0.797 (± 0.577)	1.000 (± 0.000)	1.982 (± 0.126)	0.623 (± 0.084)
		VI: DiagonalNormal	0.866 (± 0.085)	0.954 (± 1.022)	0.651 (± 0.549)	0.810 (± 0.036)	0.441 (± 0.252)	0.384 (± 0.089)
		VI: MultivariateNormal	0.765 (± 0.100)	0.537 (± 1.019)	0.633 (± 1.067)	0.711 (± 0.038)	0.148 (± 0.093)	**0.279** (± 0.056)
		VI: Structured Normal	0.758 (± 0.098)	0.447 (± 0.818)	0.572 (± 0.816)	0.705 (± 0.032)	0.140 (± 0.081)	**0.269** (± 0.045)
		VI: IAF	0.814 (± 0.105)	0.953 (± 1.165)	0.881 (± 1.067)	0.777 (± 0.106)	0.684 (± 0.939)	0.625 (± 0.525)
		ICL (ours)	**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 5	20	Laplace Approximation	**1.000** (± 0.000)	2.367 (± 0.555)	2.780 (± 1.271)	**1.000** (± 0.000)	2.200 (± 0.041)	2.444 (± 0.619)
		VI: DiagonalNormal	**0.938** (± 0.098)	**1.153** (± 0.954)	**2.552** (± 1.147)	**0.967** (± 0.012)	0.547 (± 0.233)	**1.973** (± 0.452)
		VI: MultivariateNormal	**0.929** (± 0.082)	**0.710** (± 0.768)	**2.473** (± 1.145)	**0.928** (± 0.016)	**0.250** (± 0.079)	**1.776** (± 0.399)
		VI: Structured Normal	**0.909** (± 0.082)	**0.397** (± 0.442)	**2.246** (± 1.244)	**0.924** (± 0.018)	**0.202** (± 0.094)	**1.775** (± 0.430)
		VI: IAF	**0.934** (± 0.092)	1.325 (± 1.161)	4.899 (± 1.320)	**0.980** (± 0.016)	0.892 (± 0.404)	3.593 (± 0.597)
		ICL (ours)	**0.961** (± 0.046)	1.330 (± 1.125)	5.084 (± 1.297)	**0.981** (± 0.014)	1.162 (± 0.461)	4.804 (± 0.578)
Scenario 5	50	Laplace Approximation	**1.000** (± 0.000)	2.582 (± 0.606)	**5.765** (± 1.540)	1.000 (± nan)	2.322 (± nan)	3.485 (± nan)
		VI: DiagonalNormal	**0.925** (± 0.074)	0.925 (± 1.056)	**6.461** (± 1.877)	0.972 (± nan)	0.186 (± nan)	3.251 (± nan)
		VI: MultivariateNormal	**0.934** (± 0.064)	**0.825** (± 0.972)	**6.404** (± 1.882)	0.969 (± nan)	0.165 (± nan)	3.223 (± nan)
		VI: Structured Normal	**0.927** (± 0.068)	**0.481** (± 0.588)	**6.420** (± 1.970)	0.961 (± nan)	0.072 (± nan)	3.324 (± nan)
		VI: IAF	**0.925** (± 0.069)	**0.792** (± 0.975)	8.458 (± 0.864)	0.996 (± nan)	0.519 (± nan)	4.645 (± nan)
		ICL (ours)	**0.998** (± 0.002)	**0.762** (± 0.987)	8.195 (± 0.820)	1.000 (± nan)	0.984 (± nan)	7.288 (± nan)

Table 2: Evaluating the predictive performance across 50 synthetic and 17 real-world datasets in GLM scenario 2 for different dimensionalities. All results within two standard errors of the best average result for each scenario are marked in **bold**. Due to the limitations of the number of features in the real-world data, we can only use 5 datasets for 20 and one dataset for 50 dimensions. We find that the quality of the samples by the in-context learner, when evaluated based on predictive performance, decreases consistently with an increase in the dimensionality of the problem.

Scenario	Dim.	Model	RMSE Real-World (↓)	RMSE Synthetic (↓)
Scenario 2	5	HMC	**0.559** (± 0.023)	**0.556** (± 0.049)
		Laplace Approximation	**0.561** (± 0.022)	**0.557** (± 0.049)
		VI: DiagonalNormal	**0.560** (± 0.023)	**0.557** (± 0.049)
		VI: MultivariateNormal	**0.559** (± 0.023)	**0.556** (± 0.049)
		VI: Structured Normal	**0.604** (± 0.016)	0.685 (± 0.054)
		VI: IAF	**0.563** (± 0.023)	**0.557** (± 0.049)
		ICL (ours)	**0.561** (± 0.019)	**0.653** (± 0.049)
		MAP	0.513 (± 0.023)	0.522 (± 0.048)
		TabPFN	0.449 (± 0.034)	0.498 (± 0.047)
Scenario 2	20	HMC	**0.682** (± 0.029)	**0.536** (± 0.041)
		Laplace Approximation	**0.682** (± 0.030)	**0.538** (± 0.040)
		VI: DiagonalNormal	**0.680** (± 0.029)	**0.539** (± 0.041)
		VI: MultivariateNormal	**0.685** (± 0.029)	**0.537** (± 0.041)
		VI: Structured Normal	0.746 (± 0.019)	0.681 (± 0.041)
		VI: IAF	**0.683** (± 0.029)	**0.539** (± 0.041)
		ICL (ours)	0.777 (± 0.011)	1.122 (± 0.078)
		MAP	0.578 (± 0.025)	0.472 (± 0.039)
		TabPFN	0.470 (± 0.044)	0.446 (± 0.038)
Scenario 2	50	HMC	0.669 (± nan)	**0.713** (± 0.060)
		Laplace Approximation	0.594 (± nan)	0.878 (± 0.068)
		VI: DiagonalNormal	0.582 (± nan)	0.870 (± 0.065)
		VI: MultivariateNormal	0.729 (± nan)	**0.764** (± 0.066)
		VI: Structured Normal	0.922 (± nan)	1.116 (± 0.074)
		VI: IAF	0.695 (± nan)	**0.770** (± 0.060)
		ICL (ours)	1.256 (± nan)	2.343 (± 0.230)
		MAP	0.301 (± nan)	0.398 (± 0.047)
		TabPFN	0.235 (± nan)	0.570 (± 0.053)

Table 3: Evaluating the predictive performance across 50 synthetic and 17 real-world datasets in GLM scenario 2 for different dimensionalities. All results within two standard errors of the best average result for each scenario are marked in **bold**. Due to the limitations of the number of features in the real-world data, we can only use 5 datasets for 20 and one dataset for 50 dimensions. We find that the quality of the samples by the in-context learner, when evaluated based on predictive performance, decreases consistently with an increase in the dimensionality of the problem.

Scenario	**Dim.**	**Model**	**RMSE Real-World** (↓)	**RMSE Synthetic** (↓)
Scenario 3	5	HMC	**0.684** (± 0.027)	**0.512** (± 0.040)
		Laplace Approximation	**0.688** (± 0.026)	**0.516** (± 0.040)
		VI: DiagonalNormal	**0.686** (± 0.027)	**0.513** (± 0.040)
		VI: MultivariateNormal	**0.685** (± 0.027)	**0.512** (± 0.040)
		VI: Structured Normal	**0.733** (± 0.016)	0.607 (± 0.043)
		VI: IAF	**0.686** (± 0.027)	**0.512** (± 0.040)
		ICL (ours)	**0.690** (± 0.023)	**0.588** (± 0.045)
		MAP	0.646 (± 0.028)	0.495 (± 0.039)
		TabPFN	0.556 (± 0.041)	0.462 (± 0.037)
Scenario 3	20	HMC	**1.030** (± 0.045)	**0.621** (± 0.046)
		Laplace Approximation	**1.053** (± 0.047)	0.755 (± 0.052)
		VI: DiagonalNormal	**1.035** (± 0.043)	0.734 (± 0.053)
		VI: MultivariateNormal	**1.033** (± 0.039)	**0.705** (± 0.055)
		VI: Structured Normal	**1.095** (± 0.045)	1.033 (± 0.063)
		VI: IAF	**1.026** (± 0.045)	**0.653** (± 0.047)
		ICL (ours)	1.770 (± 0.048)	2.160 (± 0.217)
		MAP	0.861 (± 0.038)	0.581 (± 0.050)
		TabPFN	0.654 (± 0.062)	0.475 (± 0.039)
Scenario 3	50	HMC	0.858 (± nan)	**0.645** (± 0.051)
		Laplace Approximation	0.866 (± nan)	0.865 (± 0.083)
		VI: DiagonalNormal	0.788 (± nan)	0.870 (± 0.084)
		VI: MultivariateNormal	0.819 (± nan)	0.778 (± 0.066)
		VI: Structured Normal	0.812 (± nan)	1.040 (± 0.103)
		VI: IAF	0.802 (± nan)	0.846 (± 0.078)
		ICL (ours)	1.686 (± nan)	3.477 (± 0.604)
		MAP	0.539 (± nan)	0.618 (± 0.054)
		TabPFN	0.322 (± nan)	0.534 (± 0.038)

Table 4: Evaluating the predictive performance across 50 synthetic and 17 real-world datasets in GLM scenario 2 for different dimensionalities. All results within two standard errors of the best average result for each scenario are marked in **bold**. Due to the limitations of the number of features in the real-world data, we can only use 5 datasets for 20 and one dataset for 50 dimensions. We find that the quality of the samples by the in-context learner, when evaluated based on predictive performance, decreases consistently with an increase in the dimensionality of the problem.

Scenario	Dim.	Model	RMSE Real-World (↓)	RMSE Synthetic (↓)
Scenario 5	5	HMC	**0.699** (± 0.022)	**0.490** (± 0.036)
		Laplace Approximation	**0.699** (± 0.022)	**0.491** (± 0.036)
		VI: DiagonalNormal	**0.702** (± 0.022)	**0.491** (± 0.036)
		VI: MultivariateNormal	**0.698** (± 0.021)	**0.491** (± 0.036)
		VI: Structured Normal	1.507 (± 0.089)	0.741 (± 0.053)
		VI: IAF	**0.699** (± 0.022)	**0.490** (± 0.036)
		ICL (ours)	0.769 (± 0.020)	0.701 (± 0.049)
		MAP	0.658 (± 0.022)	0.471 (± 0.035)
		TabPFN	0.534 (± 0.040)	0.442 (± 0.035)
Scenario 5	20	HMC	**1.527** (± 0.055)	**0.553** (± 0.044)
		Laplace Approximation	**1.585** (± 0.065)	**0.586** (± 0.043)
		VI: DiagonalNormal	**1.554** (± 0.058)	**0.586** (± 0.042)
		VI: MultivariateNormal	**1.530** (± 0.058)	**0.564** (± 0.043)
		VI: Structured Normal	2.109 (± 0.156)	1.054 (± 0.067)
		VI: IAF	**1.548** (± 0.057)	**0.562** (± 0.043)
		ICL (ours)	3.545 (± 0.288)	1.626 (± 0.140)
		MAP	1.254 (± 0.027)	0.464 (± 0.035)
		TabPFN	0.668 (± 0.064)	0.413 (± 0.032)
Scenario 5	50	HMC	1.626 (± nan)	**0.521** (± 0.028)
		Laplace Approximation	1.541 (± nan)	0.655 (± 0.040)
		VI: DiagonalNormal	1.576 (± nan)	0.639 (± 0.041)
		VI: MultivariateNormal	1.659 (± nan)	0.592 (± 0.035)
		VI: Structured Normal	2.076 (± nan)	1.018 (± 0.102)
		VI: IAF	1.706 (± nan)	0.627 (± 0.040)
		ICL (ours)	10.319 (± nan)	1.458 (± 0.193)
		MAP	1.318 (± nan)	0.416 (± 0.018)
		TabPFN	0.330 (± nan)	0.443 (± 0.024)

References

Mittal, S., Bengio, Y., Malkin, N., and Lajoie, G. In-context parametric inference: Point or distribution estimators? *arXiv preprint arXiv:2502.11617*, 2025a.

Mittal, S., Bracher, N. L., Lajoie, G., Jaini, P., and Brubaker, M. Amortized in-context bayesian posterior estimation. *arXiv preprint arXiv:2502.06601*, 2025b.