

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 ( $\downarrow$ )	MMD ( $\downarrow$ )	$\mathcal{W}_2$ ( $\downarrow$ )	C2ST ( $\downarrow$ )	MMD ( $\downarrow$ )	$\mathcal{W}_2$ ( $\downarrow$ )
Scenario 2	5	Laplace Approximation	1.000 ( $\pm$ 0.000)	4.853 ( $\pm$ 2.333)	5.770 ( $\pm$ 5.946)	1.000 ( $\pm$ 0.000)	2.572 ( $\pm$ 0.206)	0.809 ( $\pm$ 0.149)
		VI: DiagonalNormal	0.957 ( $\pm$ 0.091)	3.906 ( $\pm$ 2.679)	5.628 ( $\pm$ 6.092)	0.892 ( $\pm$ 0.044)	0.847 ( $\pm$ 0.389)	<b>0.530</b> ( $\pm$ 0.175)
		VI: MultivariateNormal	0.910 ( $\pm$ 0.131)	3.407 ( $\pm$ 2.781)	5.584 ( $\pm$ 6.104)	0.820 ( $\pm$ 0.031)	0.243 ( $\pm$ 0.148)	<b>0.408</b> ( $\pm$ 0.118)
		VI: Structured Normal	0.908 ( $\pm$ 0.119)	3.139 ( $\pm$ 2.763)	5.480 ( $\pm$ 6.164)	0.824 ( $\pm$ 0.023)	0.215 ( $\pm$ 0.110)	<b>0.392</b> ( $\pm$ 0.109)
		VI: IAF	0.968 ( $\pm$ 0.063)	4.416 ( $\pm$ 2.473)	7.474 ( $\pm$ 6.235)	0.888 ( $\pm$ 0.067)	0.921 ( $\pm$ 0.860)	0.942 ( $\pm$ 0.733)
		ICL (ours)	<b>0.839</b> ( $\pm$ 0.072)	<b>0.707</b> ( $\pm$ 0.658)	<b>1.111</b> ( $\pm$ 0.300)	<b>0.768</b> ( $\pm$ 0.033)	<b>0.143</b> ( $\pm$ 0.089)	<b>0.411</b> ( $\pm$ 0.094)
Scenario 2	20	Laplace Approximation	1.000 ( $\pm$ 0.000)	2.314 ( $\pm$ 0.237)	3.069 ( $\pm$ 1.168)	1.000 ( $\pm$ 0.000)	2.222 ( $\pm$ 0.018)	2.847 ( $\pm$ 0.305)
		VI: DiagonalNormal	0.904 ( $\pm$ 0.168)	1.292 ( $\pm$ 0.937)	<b>2.863</b> ( $\pm$ 0.919)	0.990 ( $\pm$ 0.009)	1.277 ( $\pm$ 0.452)	2.483 ( $\pm$ 0.318)
		VI: MultivariateNormal	0.851 ( $\pm$ 0.134)	<b>0.492</b> ( $\pm$ 0.547)	<b>2.694</b> ( $\pm$ 0.916)	0.843 ( $\pm$ 0.069)	0.243 ( $\pm$ 0.170)	<b>2.166</b> ( $\pm$ 0.266)
		VI: Structured Normal	<b>0.697</b> ( $\pm$ 0.065)	<b>0.070</b> ( $\pm$ 0.099)	<b>2.497</b> ( $\pm$ 0.993)	<b>0.655</b> ( $\pm$ 0.031)	<b>0.029</b> ( $\pm$ 0.025)	<b>2.191</b> ( $\pm$ 0.271)
		VI: IAF	0.916 ( $\pm$ 0.110)	1.062 ( $\pm$ 1.076)	4.191 ( $\pm$ 0.623)	0.952 ( $\pm$ 0.025)	0.515 ( $\pm$ 0.242)	3.331 ( $\pm$ 0.371)
		ICL (ours)	0.955 ( $\pm$ 0.057)	1.131 ( $\pm$ 1.035)	4.945 ( $\pm$ 0.836)	0.968 ( $\pm$ 0.020)	0.724 ( $\pm$ 0.278)	4.356 ( $\pm$ 0.302)
Scenario 2	50	Laplace Approximation	<b>1.000</b> ( $\pm$ 0.000)	2.437 ( $\pm$ 0.271)	<b>5.728</b> ( $\pm$ 1.358)	1.000 ( $\pm$ nan)	2.350 ( $\pm$ nan)	5.620 ( $\pm$ nan)
		VI: DiagonalNormal	<b>0.853</b> ( $\pm$ 0.182)	0.787 ( $\pm$ 0.687)	<b>6.224</b> ( $\pm$ 1.225)	0.996 ( $\pm$ nan)	1.080 ( $\pm$ nan)	5.426 ( $\pm$ nan)
		VI: MultivariateNormal	<b>0.878</b> ( $\pm$ 0.150)	<b>0.688</b> ( $\pm$ 0.620)	<b>6.206</b> ( $\pm$ 1.244)	0.994 ( $\pm$ nan)	0.791 ( $\pm$ nan)	5.305 ( $\pm$ nan)
		VI: Structured Normal	<b>0.865</b> ( $\pm$ 0.081)	<b>0.186</b> ( $\pm$ 0.169)	5.874 ( $\pm$ 1.233)	0.819 ( $\pm$ nan)	0.093 ( $\pm$ nan)	5.660 ( $\pm$ nan)
		VI: IAF	<b>0.909</b> ( $\pm$ 0.130)	0.649 ( $\pm$ 0.650)	7.465 ( $\pm$ 0.335)	0.985 ( $\pm$ nan)	0.426 ( $\pm$ nan)	6.426 ( $\pm$ nan)
		ICL (ours)	<b>0.972</b> ( $\pm$ 0.039)	0.741 ( $\pm$ 0.713)	8.313 ( $\pm$ 0.608)	0.971 ( $\pm$ nan)	0.405 ( $\pm$ nan)	7.718 ( $\pm$ nan)
Scenario 3	5	Laplace Approximation	1.000 ( $\pm$ 0.000)	2.203 ( $\pm$ 0.997)	1.170 ( $\pm$ 0.949)	1.000 ( $\pm$ 0.000)	1.841 ( $\pm$ 0.185)	0.729 ( $\pm$ 0.175)
		VI: DiagonalNormal	0.866 ( $\pm$ 0.101)	1.069 ( $\pm$ 1.150)	0.846 ( $\pm$ 0.747)	0.797 ( $\pm$ 0.083)	0.526 ( $\pm$ 0.361)	0.480 ( $\pm$ 0.207)
		VI: MultivariateNormal	<b>0.656</b> ( $\pm$ 0.131)	<b>0.445</b> ( $\pm$ 1.061)	<b>0.660</b> ( $\pm$ 0.737)	<b>0.560</b> ( $\pm$ 0.035)	<b>0.032</b> ( $\pm$ 0.028)	<b>0.249</b> ( $\pm$ 0.069)
		VI: Structured Normal	<b>0.653</b> ( $\pm$ 0.125)	<b>0.421</b> ( $\pm$ 0.993)	<b>0.659</b> ( $\pm$ 0.736)	<b>0.552</b> ( $\pm$ 0.028)	<b>0.027</b> ( $\pm$ 0.015)	<b>0.239</b> ( $\pm$ 0.055)
		VI: IAF	0.751 ( $\pm$ 0.148)	0.939 ( $\pm$ 1.349)	0.964 ( $\pm$ 0.924)	0.673 ( $\pm$ 0.141)	0.399 ( $\pm$ 0.543)	0.563 ( $\pm$ 0.433)
		ICL (ours)	<b>0.611</b> ( $\pm$ 0.070)	<b>0.089</b> ( $\pm$ 0.114)	<b>0.423</b> ( $\pm$ 0.348)	<b>0.576</b> ( $\pm$ 0.027)	<b>0.037</b> ( $\pm$ 0.026)	<b>0.257</b> ( $\pm$ 0.044)
Scenario 3	20	Laplace Approximation	1.000 ( $\pm$ 0.000)	2.726 ( $\pm$ 1.116)	4.127 ( $\pm$ 1.927)	1.000 ( $\pm$ 0.000)	2.234 ( $\pm$ 0.092)	3.589 ( $\pm$ 0.519)
		VI: DiagonalNormal	<b>0.912</b> ( $\pm$ 0.134)	<b>1.704</b> ( $\pm$ 1.467)	<b>3.933</b> ( $\pm$ 1.574)	0.983 ( $\pm$ 0.014)	1.298 ( $\pm$ 0.443)	3.147 ( $\pm$ 0.557)
		VI: MultivariateNormal	<b>0.863</b> ( $\pm$ 0.113)	<b>0.937</b> ( $\pm$ 1.174)	<b>3.754</b> ( $\pm$ 1.650)	<b>0.796</b> ( $\pm$ 0.099)	<b>0.268</b> ( $\pm$ 0.226)	<b>2.645</b> ( $\pm$ 0.466)
		VI: Structured Normal	<b>0.768</b> ( $\pm$ 0.109)	<b>0.302</b> ( $\pm$ 0.518)	<b>3.151</b> ( $\pm$ 1.663)	<b>0.722</b> ( $\pm$ 0.073)	<b>0.131</b> ( $\pm$ 0.141)	<b>2.579</b> ( $\pm$ 0.399)
		VI: IAF	<b>0.908</b> ( $\pm$ 0.133)	<b>1.657</b> ( $\pm$ 1.476)	5.543 ( $\pm$ 1.120)	0.936 ( $\pm$ 0.041)	0.548 ( $\pm$ 0.341)	3.678 ( $\pm$ 0.670)
		ICL (ours)	<b>0.902</b> ( $\pm$ 0.076)	<b>1.053</b> ( $\pm$ 0.782)	6.206 ( $\pm$ 0.783)	0.932 ( $\pm$ 0.019)	0.635 ( $\pm$ 0.183)	5.281 ( $\pm$ 0.317)
Scenario 3	50	Laplace Approximation	<b>1.000</b> ( $\pm$ 0.000)	2.700 ( $\pm$ 0.789)	<b>8.841</b> ( $\pm$ 1.691)	1.000 ( $\pm$ nan)	2.348 ( $\pm$ nan)	7.049 ( $\pm$ nan)
		VI: DiagonalNormal	<b>0.870</b> ( $\pm$ 0.127)	<b>1.154</b> ( $\pm$ 1.321)	<b>9.180</b> ( $\pm$ 1.513)	0.997 ( $\pm$ nan)	1.393 ( $\pm$ nan)	6.791 ( $\pm$ nan)
		VI: MultivariateNormal	<b>0.896</b> ( $\pm$ 0.101)	<b>1.027</b> ( $\pm$ 1.157)	<b>9.175</b> ( $\pm$ 1.555)	0.998 ( $\pm$ nan)	1.092 ( $\pm$ nan)	6.667 ( $\pm$ nan)
		VI: Structured Normal	<b>0.873</b> ( $\pm$ 0.112)	<b>0.539</b> ( $\pm$ 0.667)	<b>9.118</b> ( $\pm$ 1.538)	0.958 ( $\pm$ nan)	0.420 ( $\pm$ nan)	6.665 ( $\pm$ nan)
		VI: IAF	<b>0.869</b> ( $\pm$ 0.124)	<b>0.751</b> ( $\pm$ 0.939)	<b>9.917</b> ( $\pm$ 0.870)	0.971 ( $\pm$ nan)	0.417 ( $\pm$ nan)	7.411 ( $\pm$ nan)
		ICL (ours)	<b>0.931</b> ( $\pm$ 0.062)	<b>0.784</b> ( $\pm$ 0.884)	10.063 ( $\pm$ 0.930)	0.965 ( $\pm$ nan)	0.347 ( $\pm$ nan)	8.482 ( $\pm$ nan)
Scenario 5	5	Laplace Approximation	1.000 ( $\pm$ 0.000)	2.060 ( $\pm$ 0.472)	0.797 ( $\pm$ 0.577)	1.000 ( $\pm$ 0.000)	1.982 ( $\pm$ 0.126)	0.623 ( $\pm$ 0.084)
		VI: DiagonalNormal	0.866 ( $\pm$ 0.085)	0.954 ( $\pm$ 1.022)	0.651 ( $\pm$ 0.549)	0.810 ( $\pm$ 0.036)	0.441 ( $\pm$ 0.252)	0.384 ( $\pm$ 0.089)
		VI: MultivariateNormal	0.765 ( $\pm$ 0.100)	0.537 ( $\pm$ 1.019)	0.633 ( $\pm$ 1.067)	0.711 ( $\pm$ 0.038)	0.148 ( $\pm$ 0.093)	<b>0.279</b> ( $\pm$ 0.056)
		VI: Structured Normal	0.758 ( $\pm$ 0.098)	0.447 ( $\pm$ 0.818)	0.572 ( $\pm$ 0.816)	0.705 ( $\pm$ 0.032)	0.140 ( $\pm$ 0.081)	<b>0.269</b> ( $\pm$ 0.045)
		VI: IAF	0.814 ( $\pm$ 0.105)	0.953 ( $\pm$ 1.165)	0.881 ( $\pm$ 1.067)	0.777 ( $\pm$ 0.106)	0.684 ( $\pm$ 0.939)	0.625 ( $\pm$ 0.525)
		ICL (ours)	<b>0.621</b> ( $\pm$ 0.063)	<b>0.067</b> ( $\pm$ 0.080)	<b>0.299</b> ( $\pm$ 0.195)	<b>0.610</b> ( $\pm$ 0.045)	<b>0.046</b> ( $\pm$ 0.020)	<b>0.242</b> ( $\pm$ 0.038)
Scenario 5	20	Laplace Approximation	<b>1.000</b> ( $\pm$ 0.000)	2.367 ( $\pm$ 0.555)	2.780 ( $\pm$ 1.271)	<b>1.000</b> ( $\pm$ 0.000)	2.200 ( $\pm$ 0.041)	2.444 ( $\pm$ 0.619)
		VI: DiagonalNormal	<b>0.938</b> ( $\pm$ 0.098)	<b>1.153</b> ( $\pm$ 0.954)	<b>2.552</b> ( $\pm$ 1.147)	<b>0.967</b> ( $\pm$ 0.012)	0.547 ( $\pm$ 0.233)	<b>1.973</b> ( $\pm$ 0.452)
		VI: MultivariateNormal	<b>0.929</b> ( $\pm$ 0.082)	<b>0.710</b> ( $\pm$ 0.768)	<b>2.473</b> ( $\pm$ 1.145)	<b>0.928</b> ( $\pm$ 0.016)	<b>0.250</b> ( $\pm$ 0.079)	<b>1.776</b> ( $\pm$ 0.399)
		VI: Structured Normal	<b>0.909</b> ( $\pm$ 0.082)	<b>0.397</b> ( $\pm$ 0.442)	<b>2.246</b> ( $\pm$ 1.244)	<b>0.924</b> ( $\pm$ 0.018)	<b>0.202</b> ( $\pm$ 0.094)	<b>1.775</b> ( $\pm$ 0.430)
		VI: IAF	<b>0.934</b> ( $\pm$ 0.092)	1.325 ( $\pm$ 1.161)	4.899 ( $\pm$ 1.320)	<b>0.980</b> ( $\pm$ 0.016)	0.892 ( $\pm$ 0.404)	3.593 ( $\pm$ 0.597)
		ICL (ours)	<b>0.961</b> ( $\pm$ 0.046)	1.330 ( $\pm$ 1.125)	5.084 ( $\pm$ 1.297)	<b>0.981</b> ( $\pm$ 0.014)	1.162 ( $\pm$ 0.461)	4.804 ( $\pm$ 0.578)
Scenario 5	50	Laplace Approximation	<b>1.000</b> ( $\pm$ 0.000)	2.582 ( $\pm$ 0.606)	<b>5.765</b> ( $\pm$ 1.540)	1.000 ( $\pm$ nan)	2.322 ( $\pm$ nan)	3.485 ( $\pm$ nan)
		VI: DiagonalNormal	<b>0.925</b> ( $\pm$ 0.074)	0.925 ( $\pm$ 1.056)	<b>6.461</b> ( $\pm$ 1.877)	0.972 ( $\pm$ nan)	0.186 ( $\pm$ nan)	3.251 ( $\pm$ nan)
		VI: MultivariateNormal	<b>0.934</b> ( $\pm$ 0.064)	<b>0.825</b> ( $\pm$ 0.972)	<b>6.404</b> ( $\pm$ 1.882)	0.969 ( $\pm$ nan)	0.165 ( $\pm$ nan)	3.223 ( $\pm$ nan)
		VI: Structured Normal	<b>0.927</b> ( $\pm$ 0.068)	<b>0.481</b> ( $\pm$ 0.588)	<b>6.420</b> ( $\pm$ 1.970)	0.961 ( $\pm$ nan)	0.072 ( $\pm$ nan)	3.324 ( $\pm$ nan)
		VI: IAF	<b>0.925</b> ( $\pm$ 0.069)	<b>0.792</b> ( $\pm$ 0.975)	8.458 ( $\pm$ 0.864)	0.996 ( $\pm$ nan)	0.519 ( $\pm$ nan)	4.645 ( $\pm$ nan)
		ICL (ours)	<b>0.998</b> ( $\pm$ 0.002)	<b>0.762</b> ( $\pm$ 0.987)	8.195 ( $\pm$ 0.820)	1.000 ( $\pm$ nan)	0.984 ( $\pm$ nan)	7.288 ( $\pm$ 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 ( $\downarrow$ )	RMSE Synthetic ( $\downarrow$ )
Scenario 2	5	HMC	<b>0.559</b> ( $\pm 0.023$ )	<b>0.556</b> ( $\pm 0.049$ )
		Laplace Approximation	<b>0.561</b> ( $\pm 0.022$ )	<b>0.557</b> ( $\pm 0.049$ )
		VI: DiagonalNormal	<b>0.560</b> ( $\pm 0.023$ )	<b>0.557</b> ( $\pm 0.049$ )
		VI: MultivariateNormal	<b>0.559</b> ( $\pm 0.023$ )	<b>0.556</b> ( $\pm 0.049$ )
		VI: Structured Normal	<b>0.604</b> ( $\pm 0.016$ )	0.685 ( $\pm 0.054$ )
		VI: IAF	<b>0.563</b> ( $\pm 0.023$ )	<b>0.557</b> ( $\pm 0.049$ )
		ICL (ours)	<b>0.561</b> ( $\pm 0.019$ )	<b>0.653</b> ( $\pm 0.049$ )
		MAP	0.513 ( $\pm 0.023$ )	0.522 ( $\pm 0.048$ )
		TabPFN	0.449 ( $\pm 0.034$ )	0.498 ( $\pm 0.047$ )
Scenario 2	20	HMC	<b>0.682</b> ( $\pm 0.029$ )	<b>0.536</b> ( $\pm 0.041$ )
		Laplace Approximation	<b>0.682</b> ( $\pm 0.030$ )	<b>0.538</b> ( $\pm 0.040$ )
		VI: DiagonalNormal	<b>0.680</b> ( $\pm 0.029$ )	<b>0.539</b> ( $\pm 0.041$ )
		VI: MultivariateNormal	<b>0.685</b> ( $\pm 0.029$ )	<b>0.537</b> ( $\pm 0.041$ )
		VI: Structured Normal	0.746 ( $\pm 0.019$ )	0.681 ( $\pm 0.041$ )
		VI: IAF	<b>0.683</b> ( $\pm 0.029$ )	<b>0.539</b> ( $\pm 0.041$ )
		ICL (ours)	0.777 ( $\pm 0.011$ )	1.122 ( $\pm 0.078$ )
		MAP	0.578 ( $\pm 0.025$ )	0.472 ( $\pm 0.039$ )
		TabPFN	0.470 ( $\pm 0.044$ )	0.446 ( $\pm 0.038$ )
Scenario 2	50	HMC	0.669 ( $\pm \text{nan}$ )	<b>0.713</b> ( $\pm 0.060$ )
		Laplace Approximation	0.594 ( $\pm \text{nan}$ )	0.878 ( $\pm 0.068$ )
		VI: DiagonalNormal	0.582 ( $\pm \text{nan}$ )	0.870 ( $\pm 0.065$ )
		VI: MultivariateNormal	0.729 ( $\pm \text{nan}$ )	<b>0.764</b> ( $\pm 0.066$ )
		VI: Structured Normal	0.922 ( $\pm \text{nan}$ )	1.116 ( $\pm 0.074$ )
		VI: IAF	0.695 ( $\pm \text{nan}$ )	<b>0.770</b> ( $\pm 0.060$ )
		ICL (ours)	1.256 ( $\pm \text{nan}$ )	2.343 ( $\pm 0.230$ )
		MAP	0.301 ( $\pm \text{nan}$ )	0.398 ( $\pm 0.047$ )
		TabPFN	0.235 ( $\pm \text{nan}$ )	0.570 ( $\pm 0.053$ )

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

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

## 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.