

Table 1: Generalized Linear Models: Ablation with respect to the dimensionality of the problem on 50 synthetic and 17 real-world datasets for scenarios 1,2 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. Note that the randomness in the results, especially for higher dimensionalities, can in rare cases lead to better mean values. This is most likely not significant when taking the standard error into account.

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 1	5	Laplace Approximation	1.000 (\pm 0.000)	2.738 (\pm 0.721)	0.825 (\pm 0.279)	1.000 (\pm 0.000)	2.150 (\pm 0.323)	0.642 (\pm 0.124)
		VI: DiagonalNormal	0.904 (\pm 0.076)	1.452 (\pm 0.984)	0.669 (\pm 0.301)	0.797 (\pm 0.083)	0.612 (\pm 0.511)	0.414 (\pm 0.152)
		VI: MultivariateNormal	0.750 (\pm 0.128)	0.735 (\pm 0.733)	0.565 (\pm 0.292)	0.607 (\pm 0.070)	0.167 (\pm 0.196)	0.301 (\pm 0.123)
		VI: Structured Normal	0.753 (\pm 0.126)	0.736 (\pm 0.737)	0.570 (\pm 0.310)	0.600 (\pm 0.070)	0.169 (\pm 0.214)	0.306 (\pm 0.131)
		VI: IAF	0.777 (\pm 0.122)	0.864 (\pm 0.844)	0.725 (\pm 0.523)	0.683 (\pm 0.132)	0.440 (\pm 0.559)	0.503 (\pm 0.383)
		HMC	0.745 (\pm 0.130)	0.722 (\pm 0.732)	0.569 (\pm 0.301)	0.595 (\pm 0.075)	0.173 (\pm 0.213)	0.321 (\pm 0.140)
		ICL (ours)	0.765 (\pm 0.123)	0.767 (\pm 0.727)	0.585 (\pm 0.301)	0.614 (\pm 0.074)	0.175 (\pm 0.219)	0.310 (\pm 0.138)
Scenario 1	20	Laplace Approximation	1.000 (\pm 0.000)	2.237 (\pm 0.024)	3.252 (\pm 1.172)	1.000 (\pm 0.000)	2.206 (\pm 0.021)	2.792 (\pm 0.339)
		VI: DiagonalNormal	0.843 (\pm 0.204)	1.056 (\pm 0.869)	2.976 (\pm 0.927)	0.983 (\pm 0.019)	1.217 (\pm 0.463)	2.406 (\pm 0.348)
		VI: MultivariateNormal	0.789 (\pm 0.140)	0.714 (\pm 0.351)	2.774 (\pm 0.995)	0.768 (\pm 0.120)	0.180 (\pm 0.159)	2.064 (\pm 0.306)
		VI: Structured Normal	0.792 (\pm 0.109)	0.708 (\pm 0.153)	2.703 (\pm 1.069)	0.668 (\pm 0.080)	0.168 (\pm 0.071)	2.052 (\pm 0.275)
		VI: IAF	0.832 (\pm 0.196)	0.847 (\pm 1.016)	4.015 (\pm 0.415)	0.939 (\pm 0.024)	0.508 (\pm 0.200)	3.140 (\pm 0.290)
		ICL (ours)	0.849 (\pm 0.171)	0.844 (\pm 0.905)	4.564 (\pm 0.622)	0.970 (\pm 0.020)	0.724 (\pm 0.287)	4.250 (\pm 0.312)
		ICL (ours)	0.849 (\pm 0.171)	0.844 (\pm 0.905)	4.564 (\pm 0.622)	0.970 (\pm 0.020)	0.724 (\pm 0.287)	4.250 (\pm 0.312)
Scenario 1	50	Laplace Approximation	1.000 (\pm 0.000)	2.401 (\pm 0.282)	5.152 (\pm 2.268)	1.000 (\pm nan)	2.339 (\pm nan)	6.642 (\pm nan)
		VI: DiagonalNormal	0.812 (\pm 0.197)	0.956 (\pm 1.016)	5.541 (\pm 2.356)	0.915 (\pm nan)	0.508 (\pm nan)	6.200 (\pm nan)
		VI: MultivariateNormal	0.839 (\pm 0.148)	0.926 (\pm 0.682)	5.514 (\pm 2.370)	0.905 (\pm nan)	0.790 (\pm nan)	6.258 (\pm nan)
		VI: Structured Normal	0.823 (\pm 0.160)	0.844 (\pm 0.480)	5.752 (\pm 2.098)	0.910 (\pm nan)	1.122 (\pm nan)	6.898 (\pm nan)
		VI: IAF	0.820 (\pm 0.182)	0.814 (\pm 0.987)	6.696 (\pm 1.207)	0.938 (\pm nan)	0.256 (\pm nan)	6.869 (\pm nan)
		ICL (ours)	0.787 (\pm 0.217)	1.015 (\pm 1.255)	8.278 (\pm 0.821)	0.979 (\pm nan)	0.413 (\pm nan)	8.368 (\pm nan)
		ICL (ours)	0.787 (\pm 0.217)	1.015 (\pm 1.255)	8.278 (\pm 0.821)	0.979 (\pm nan)	0.413 (\pm nan)	8.368 (\pm nan)
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)	0.530 (\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)	0.408 (\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)	0.392 (\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)	0.839 (\pm 0.072)	0.707 (\pm 0.658)	1.111 (\pm 0.300)	0.768 (\pm 0.033)	0.143 (\pm 0.089)	0.411 (\pm 0.094)
		ICL (ours)	0.839 (\pm 0.072)	0.707 (\pm 0.658)	1.111 (\pm 0.300)	0.768 (\pm 0.033)	0.143 (\pm 0.089)	0.411 (\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)	2.863 (\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)	0.492 (\pm 0.547)	2.694 (\pm 0.916)	0.843 (\pm 0.069)	0.243 (\pm 0.170)	2.166 (\pm 0.266)
		VI: Structured Normal	0.697 (\pm 0.065)	0.070 (\pm 0.099)	2.497 (\pm 0.993)	0.655 (\pm 0.031)	0.029 (\pm 0.025)	2.191 (\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)
		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	1.000 (\pm 0.000)	2.437 (\pm 0.271)	5.728 (\pm 1.358)	1.000 (\pm nan)	2.350 (\pm nan)	5.620 (\pm nan)
		VI: DiagonalNormal	0.853 (\pm 0.182)	0.787 (\pm 0.687)	6.224 (\pm 1.225)	0.996 (\pm nan)	1.080 (\pm nan)	5.426 (\pm nan)
		VI: MultivariateNormal	0.878 (\pm 0.150)	0.688 (\pm 0.620)	6.206 (\pm 1.244)	0.994 (\pm nan)	0.791 (\pm nan)	5.305 (\pm nan)
		VI: Structured Normal	0.865 (\pm 0.081)	0.186 (\pm 0.169)	5.874 (\pm 1.233)	0.819 (\pm nan)	0.093 (\pm nan)	5.660 (\pm nan)
		VI: IAF	0.909 (\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)	0.972 (\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)
		ICL (ours)	0.972 (\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	0.656 (\pm 0.131)	0.445 (\pm 1.061)	0.660 (\pm 0.737)	0.560 (\pm 0.035)	0.032 (\pm 0.028)	0.249 (\pm 0.069)
		VI: Structured Normal	0.653 (\pm 0.125)	0.421 (\pm 0.993)	0.659 (\pm 0.736)	0.552 (\pm 0.028)	0.027 (\pm 0.015)	0.239 (\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)	0.611 (\pm 0.070)	0.089 (\pm 0.114)	0.423 (\pm 0.348)	0.576 (\pm 0.027)	0.037 (\pm 0.026)	0.257 (\pm 0.044)
		ICL (ours)	0.611 (\pm 0.070)	0.089 (\pm 0.114)	0.423 (\pm 0.348)	0.576 (\pm 0.027)	0.037 (\pm 0.026)	0.257 (\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	0.912 (\pm 0.134)	1.704 (\pm 1.467)	3.933 (\pm 1.574)	0.983 (\pm 0.014)	1.298 (\pm 0.443)	3.147 (\pm 0.557)
		VI: MultivariateNormal	0.863 (\pm 0.113)	0.937 (\pm 1.174)	3.754 (\pm 1.650)	0.796 (\pm 0.099)	0.268 (\pm 0.226)	2.645 (\pm 0.466)
		VI: Structured Normal	0.768 (\pm 0.109)	0.302 (\pm 0.518)	3.151 (\pm 1.663)	0.722 (\pm 0.073)	0.131 (\pm 0.141)	2.579 (\pm 0.399)
		VI: IAF	0.908 (\pm 0.133)	1.657 (\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)	0.902 (\pm 0.076)	1.053 (\pm 0.782)	6.206 (\pm 0.783)	0.932 (\pm 0.019)	0.635 (\pm 0.183)	5.281 (\pm 0.317)
		ICL (ours)	0.902 (\pm 0.076)	1.053 (\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	1.000 (\pm 0.000)	2.700 (\pm 0.789)	8.841 (\pm 1.691)	1.000 (\pm nan)	2.348 (\pm nan)	7.049 (\pm nan)
		VI: DiagonalNormal	0.870 (\pm 0.127)	1.154 (\pm 1.321)	9.180 (\pm 1.513)	0.997 (\pm nan)	1.393 (\pm nan)	6.791 (\pm nan)
		VI: MultivariateNormal	0.896 (\pm 0.101)	1.027 (\pm 1.157)	9.175 (\pm 1.555)	0.998 (\pm nan)	1.092 (\pm nan)	6.667 (\pm nan)
		VI: Structured Normal	0.873 (\pm 0.112)	0.539 (\pm 0.667)	9.118 (\pm 1.538)	0.958 (\pm nan)	0.420 (\pm nan)	6.665 (\pm nan)
		VI: IAF	0.869 (\pm 0.124)	0.751 (\pm 0.939)	9.917 (\pm 0.870)	0.971 (\pm nan)	0.417 (\pm nan)	7.411 (\pm nan)
		ICL (ours)	0.931 (\pm 0.062)	0.784 (\pm 0.884)	10.063 (\pm 0.930)	0.965 (\pm nan)	0.347 (\pm nan)	8.482 (\pm nan)
		ICL (ours)	0.931 (\pm 0.062)	0.784 (\pm 0.884)	10.063 (\pm 0.930)	0.965 (\pm nan)	0.347 (\pm nan)	8.482 (\pm nan)

Table 2: Generalized Linear Models: Ablation with respect to the dimensionality of the problem on 50 synthetic and 17 real-world datasets for scenarios 4, 5 and 6. 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. Similar to scenarios 1,2 and 3, 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. For 50 dimensions we find that in many cases the used metrics do not allow to significantly discriminate the performance of the different approaches. Note that the randomness in the results, especially for higher dimensionalities, can in rare cases lead to better mean values. This is most likely not significant when taking the standard error into account.

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 4	5	Laplace Approximation	1.000 (\pm 0.000)	3.511 (\pm 2.025)	2.166 (\pm 1.722)	1.000 (\pm 0.000)	2.011 (\pm 0.058)	0.993 (\pm 0.144)
		VI: DiagonalNormal	0.968 (\pm 0.036)	2.798 (\pm 2.255)	2.065 (\pm 1.745)	0.916 (\pm 0.040)	0.928 (\pm 0.339)	0.732 (\pm 0.181)
		VI: MultivariateNormal	0.855 (\pm 0.123)	1.648 (\pm 2.052)	1.853 (\pm 1.745)	0.771 (\pm 0.017)	0.087 (\pm 0.030)	0.539 (\pm 0.070)
		VI: Structured Normal	0.847 (\pm 0.116)	1.505 (\pm 1.978)	1.889 (\pm 1.883)	0.769 (\pm 0.012)	0.083 (\pm 0.018)	0.543 (\pm 0.070)
		VI: IAF	0.942 (\pm 0.077)	3.029 (\pm 2.210)	3.554 (\pm 2.715)	0.833 (\pm 0.069)	0.636 (\pm 0.756)	0.978 (\pm 0.600)
		ICL (ours)	0.753 (\pm 0.049)	0.171 (\pm 0.153)	0.631 (\pm 0.294)	0.762 (\pm 0.015)	0.105 (\pm 0.046)	0.597 (\pm 0.104)
Scenario 4	20	Laplace Approximation	1.000 (\pm 0.000)	4.929 (\pm 1.611)	8.863 (\pm 3.796)	1.000 (\pm 0.000)	3.196 (\pm 0.841)	5.186 (\pm 1.533)
		VI: DiagonalNormal	0.988 (\pm 0.060)	4.418 (\pm 2.013)	9.364 (\pm 4.281)	0.997 (\pm 0.007)	3.095 (\pm 1.417)	6.098 (\pm 2.435)
		VI: MultivariateNormal	0.986 (\pm 0.054)	3.388 (\pm 1.907)	7.910 (\pm 4.070)	0.893 (\pm 0.087)	0.534 (\pm 0.469)	3.175 (\pm 0.751)
		VI: Structured Normal	0.954 (\pm 0.076)	2.254 (\pm 1.515)	7.475 (\pm 4.224)	0.727 (\pm 0.034)	0.074 (\pm 0.070)	2.877 (\pm 0.379)
		VI: IAF	0.987 (\pm 0.059)	3.258 (\pm 1.415)	9.865 (\pm 3.515)	0.955 (\pm 0.030)	0.629 (\pm 0.308)	4.098 (\pm 0.341)
		ICL (ours)	0.978 (\pm 0.038)	1.185 (\pm 0.720)	11.335 (\pm 1.378)	0.972 (\pm 0.018)	0.668 (\pm 0.199)	9.937 (\pm 0.466)
Scenario 4	50	Laplace Approximation	1.000 (\pm 0.000)	6.695 (\pm 1.329)	12.323 (\pm 4.091)	1.000 (\pm nan)	5.491 (\pm nan)	7.518 (\pm nan)
		VI: DiagonalNormal	0.965 (\pm 0.084)	2.395 (\pm 1.958)	12.022 (\pm 3.673)	0.996 (\pm nan)	4.368 (\pm nan)	6.951 (\pm nan)
		VI: MultivariateNormal	0.984 (\pm 0.054)	5.395 (\pm 1.847)	12.141 (\pm 3.079)	1.000 (\pm nan)	5.146 (\pm nan)	9.002 (\pm nan)
		VI: Structured Normal	0.982 (\pm 0.026)	4.261 (\pm 1.191)	11.126 (\pm 3.396)	0.869 (\pm nan)	3.181 (\pm nan)	7.065 (\pm nan)
		VI: IAF	0.981 (\pm 0.048)	4.609 (\pm 1.412)	12.567 (\pm 3.131)	0.988 (\pm nan)	3.558 (\pm nan)	7.849 (\pm nan)
		ICL (ours)	0.960 (\pm 0.045)	3.792 (\pm 0.758)	14.071 (\pm 0.894)	0.974 (\pm nan)	3.443 (\pm nan)	12.546 (\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)	0.279 (\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)	0.269 (\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)	0.621 (\pm 0.063)	0.067 (\pm 0.080)	0.299 (\pm 0.195)	0.610 (\pm 0.045)	0.046 (\pm 0.020)	0.242 (\pm 0.038)
Scenario 5	20	Laplace Approximation	1.000 (\pm 0.000)	2.367 (\pm 0.555)	2.780 (\pm 1.271)	1.000 (\pm 0.000)	2.200 (\pm 0.041)	2.444 (\pm 0.619)
		VI: DiagonalNormal	0.938 (\pm 0.098)	1.153 (\pm 0.954)	2.552 (\pm 1.147)	0.967 (\pm 0.012)	0.547 (\pm 0.233)	1.973 (\pm 0.452)
		VI: MultivariateNormal	0.929 (\pm 0.082)	0.710 (\pm 0.768)	2.473 (\pm 1.145)	0.928 (\pm 0.016)	0.250 (\pm 0.079)	1.776 (\pm 0.399)
		VI: Structured Normal	0.909 (\pm 0.082)	0.397 (\pm 0.442)	2.246 (\pm 1.244)	0.924 (\pm 0.018)	0.202 (\pm 0.094)	1.775 (\pm 0.430)
		VI: IAF	0.934 (\pm 0.092)	1.325 (\pm 1.161)	4.899 (\pm 1.320)	0.980 (\pm 0.016)	0.892 (\pm 0.404)	3.593 (\pm 0.597)
		ICL (ours)	0.961 (\pm 0.046)	1.330 (\pm 1.125)	5.084 (\pm 1.297)	0.981 (\pm 0.014)	1.162 (\pm 0.461)	4.804 (\pm 0.578)
Scenario 5	50	Laplace Approximation	1.000 (\pm 0.000)	2.582 (\pm 0.606)	5.765 (\pm 1.540)	1.000 (\pm nan)	2.322 (\pm nan)	3.485 (\pm nan)
		VI: DiagonalNormal	0.925 (\pm 0.074)	0.925 (\pm 1.056)	6.461 (\pm 1.877)	0.972 (\pm nan)	0.186 (\pm nan)	3.251 (\pm nan)
		VI: MultivariateNormal	0.934 (\pm 0.064)	0.825 (\pm 0.972)	6.404 (\pm 1.882)	0.969 (\pm nan)	0.165 (\pm nan)	3.223 (\pm nan)
		VI: Structured Normal	0.927 (\pm 0.068)	0.481 (\pm 0.588)	6.420 (\pm 1.970)	0.961 (\pm nan)	0.072 (\pm nan)	3.324 (\pm nan)
		VI: IAF	0.925 (\pm 0.069)	0.792 (\pm 0.975)	8.458 (\pm 0.864)	0.996 (\pm nan)	0.519 (\pm nan)	4.645 (\pm nan)
		ICL (ours)	0.998 (\pm 0.002)	0.762 (\pm 0.987)	8.195 (\pm 0.820)	1.000 (\pm nan)	0.984 (\pm nan)	7.288 (\pm nan)
Scenario 6	5	Laplace Approximation	1.000 (\pm 0.000)	2.026 (\pm 0.027)	1.612 (\pm 0.162)	1.000 (\pm 0.000)	1.993 (\pm 0.032)	1.299 (\pm 0.106)
		VI: DiagonalNormal	0.724 (\pm 0.060)	0.185 (\pm 0.082)	0.787 (\pm 0.078)	0.703 (\pm 0.039)	0.147 (\pm 0.063)	0.637 (\pm 0.089)
		VI: MultivariateNormal	0.534 (\pm 0.018)	0.014 (\pm 0.006)	0.581 (\pm 0.074)	0.538 (\pm 0.019)	0.016 (\pm 0.007)	0.466 (\pm 0.029)
		VI: Structured Normal	0.536 (\pm 0.016)	0.014 (\pm 0.005)	0.583 (\pm 0.071)	0.536 (\pm 0.019)	0.017 (\pm 0.009)	0.469 (\pm 0.033)
		VI: IAF	0.542 (\pm 0.026)	0.031 (\pm 0.031)	0.613 (\pm 0.092)	0.535 (\pm 0.015)	0.015 (\pm 0.006)	0.467 (\pm 0.031)
		ICL (ours)	0.532 (\pm 0.019)	0.016 (\pm 0.008)	0.590 (\pm 0.066)	0.556 (\pm 0.017)	0.035 (\pm 0.015)	0.504 (\pm 0.038)
Scenario 6	20	Laplace Approximation	1.000 (\pm 0.000)	2.247 (\pm 0.006)	4.158 (\pm 0.243)	1.000 (\pm 0.000)	2.240 (\pm 0.007)	3.714 (\pm 0.127)
		VI: DiagonalNormal	0.747 (\pm 0.138)	0.136 (\pm 0.123)	3.460 (\pm 0.361)	0.836 (\pm 0.053)	0.203 (\pm 0.086)	2.977 (\pm 0.112)
		VI: MultivariateNormal	0.621 (\pm 0.016)	0.016 (\pm 0.002)	3.564 (\pm 0.290)	0.608 (\pm 0.017)	0.015 (\pm 0.003)	3.101 (\pm 0.115)
		VI: Structured Normal	0.599 (\pm 0.015)	0.012 (\pm 0.002)	3.592 (\pm 0.267)	0.584 (\pm 0.028)	0.012 (\pm 0.002)	3.120 (\pm 0.107)
		VI: IAF	0.625 (\pm 0.040)	0.019 (\pm 0.009)	3.572 (\pm 0.266)	0.636 (\pm 0.021)	0.020 (\pm 0.005)	3.106 (\pm 0.128)
		ICL (ours)	0.747 (\pm 0.148)	0.163 (\pm 0.144)	4.063 (\pm 0.184)	0.928 (\pm 0.030)	0.463 (\pm 0.162)	4.425 (\pm 0.314)
Scenario 6	50	Laplace Approximation	1.000 (\pm 0.000)	2.291 (\pm 0.003)	6.742 (\pm 0.362)	1.000 (\pm nan)	2.293 (\pm nan)	6.587 (\pm nan)
		VI: DiagonalNormal	0.761 (\pm 0.138)	0.087 (\pm 0.083)	6.909 (\pm 0.743)	0.905 (\pm nan)	0.175 (\pm nan)	6.403 (\pm nan)
		VI: MultivariateNormal	0.797 (\pm 0.100)	0.069 (\pm 0.055)	6.956 (\pm 0.736)	0.891 (\pm nan)	0.110 (\pm nan)	6.473 (\pm nan)
		VI: Structured Normal	0.647 (\pm 0.017)	0.013 (\pm 0.002)	7.218 (\pm 0.506)	0.654 (\pm nan)	0.013 (\pm nan)	6.890 (\pm nan)
		VI: IAF	0.639 (\pm 0.038)	0.014 (\pm 0.006)	7.204 (\pm 0.463)	0.692 (\pm nan)	0.024 (\pm nan)	6.887 (\pm nan)
		ICL (ours)	0.742 (\pm 0.178)	0.115 (\pm 0.124)	7.713 (\pm 0.120)	0.935 (\pm nan)	0.203 (\pm nan)	7.846 (\pm nan)

Table 3: Generalized Linear Models: Ablation with respect to the dimensionality of the problem on 50 synthetic and 17 real-world datasets for scenario 7. 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. Also for this scenario, we find that the advantages of the in-context learning approach to deteriorate for higher dimensionalities. However, in this specific scenario the in-context learner is not significantly different from the VI methods in terms of C2ST and MMD for 20 dimensions. For 50 dimensions we find that the VI method using IAF performs well, together with the in-context learning approach in terms of MMD while the C2ST score does not indicate a clear winner and \mathcal{W}_2 favors the other methods. Note that the randomness in the results, especially for higher dimensionalities, can in rare cases lead to better mean values. This is most likely not significant when taking the standard error into account.

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 7	5	Laplace Approximation	1.000 (\pm 0.000)	3.559 (\pm 1.933)	1.347 (\pm 1.067)	1.000 (\pm 0.000)	2.016 (\pm 0.080)	0.763 (\pm 0.174)
		VI: DiagonalNormal	0.938 (\pm 0.074)	2.536 (\pm 2.097)	1.142 (\pm 0.993)	0.936 (\pm 0.024)	1.029 (\pm 0.255)	0.579 (\pm 0.181)
		VI: MultivariateNormal	0.814 (\pm 0.181)	1.999 (\pm 2.283)	1.033 (\pm 0.969)	0.741 (\pm 0.020)	0.093 (\pm 0.025)	0.391 (\pm 0.074)
		VI: Structured Normal	0.824 (\pm 0.177)	1.891 (\pm 2.127)	1.041 (\pm 0.934)	0.734 (\pm 0.025)	0.072 (\pm 0.019)	0.385 (\pm 0.065)
		VI: IAF	0.939 (\pm 0.091)	2.707 (\pm 1.712)	1.590 (\pm 0.820)	0.864 (\pm 0.093)	0.830 (\pm 0.697)	1.064 (\pm 0.616)
		ICL (ours)	0.700 (\pm 0.116)	0.317 (\pm 0.355)	0.400 (\pm 0.286)	0.773 (\pm 0.048)	0.294 (\pm 0.457)	0.559 (\pm 0.256)
Scenario 7	20	Laplace Approximation	1.000 (\pm 0.000)	3.581 (\pm 2.147)	3.365 (\pm 1.583)	1.000 (\pm 0.000)	2.213 (\pm 0.024)	2.539 (\pm 0.378)
		VI: DiagonalNormal	0.887 (\pm 0.184)	2.819 (\pm 2.732)	3.637 (\pm 1.371)	0.996 (\pm 0.005)	1.734 (\pm 0.314)	2.348 (\pm 0.423)
		VI: MultivariateNormal	0.881 (\pm 0.164)	2.265 (\pm 2.573)	3.524 (\pm 1.392)	0.916 (\pm 0.085)	0.766 (\pm 0.535)	2.043 (\pm 0.516)
		VI: Structured Normal	0.850 (\pm 0.162)	1.667 (\pm 2.266)	3.186 (\pm 1.315)	0.849 (\pm 0.105)	0.391 (\pm 0.244)	1.880 (\pm 0.367)
		VI: IAF	0.867 (\pm 0.184)	1.629 (\pm 1.584)	4.875 (\pm 1.239)	0.986 (\pm 0.007)	0.895 (\pm 0.361)	4.096 (\pm 0.319)
		ICL (ours)	0.867 (\pm 0.185)	1.428 (\pm 1.352)	4.836 (\pm 1.032)	0.982 (\pm 0.010)	0.820 (\pm 0.324)	4.177 (\pm 0.368)
Scenario 7	50	Laplace Approximation	1.000 (\pm 0.000)	4.768 (\pm 1.171)	6.573 (\pm 1.038)	1.000 (\pm nan)	2.312 (\pm nan)	5.270 (\pm nan)
		VI: DiagonalNormal	0.771 (\pm 0.191)	3.263 (\pm 1.853)	6.919 (\pm 1.257)	1.000 (\pm nan)	2.237 (\pm nan)	5.417 (\pm nan)
		VI: MultivariateNormal	0.816 (\pm 0.154)	3.245 (\pm 1.793)	6.978 (\pm 1.226)	0.997 (\pm nan)	2.117 (\pm nan)	5.781 (\pm nan)
		VI: Structured Normal	0.795 (\pm 0.171)	3.126 (\pm 1.677)	6.918 (\pm 1.260)	1.000 (\pm nan)	1.879 (\pm nan)	5.461 (\pm nan)
		VI: IAF	0.769 (\pm 0.189)	2.534 (\pm 0.894)	7.895 (\pm 0.843)	0.994 (\pm nan)	0.584 (\pm nan)	7.626 (\pm nan)
		ICL (ours)	0.732 (\pm 0.216)	2.451 (\pm 0.790)	7.787 (\pm 0.661)	0.980 (\pm nan)	0.411 (\pm nan)	7.461 (\pm nan)

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. Note that the randomness in the results, especially for higher dimensionalities, can in rare cases lead to better mean values. This is most likely not significant when taking the standard error into account.

Scenario	Dim.	Model	RMSE Real-World (\downarrow)	RMSE Synthetic (\downarrow)
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 ($\pm \text{nan}$)	0.713 (± 0.060)
		Laplace Approximation	0.594 ($\pm \text{nan}$)	0.878 (± 0.068)
		VI: DiagonalNormal	0.582 ($\pm \text{nan}$)	0.870 (± 0.065)
		VI: MultivariateNormal	0.729 ($\pm \text{nan}$)	0.764 (± 0.066)
		VI: Structured Normal	0.922 ($\pm \text{nan}$)	1.116 (± 0.074)
		VI: IAF	0.695 ($\pm \text{nan}$)	0.770 (± 0.060)
		ICL (ours)	1.256 ($\pm \text{nan}$)	2.343 (± 0.230)
		MAP	0.301 ($\pm \text{nan}$)	0.398 (± 0.047)
		TabPFN	0.235 ($\pm \text{nan}$)	0.570 (± 0.053)

Table 5: 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	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 ($\pm \text{nan}$)	0.645 (± 0.051)
		Laplace Approximation	0.866 ($\pm \text{nan}$)	0.865 (± 0.083)
		VI: DiagonalNormal	0.788 ($\pm \text{nan}$)	0.870 (± 0.084)
		VI: MultivariateNormal	0.819 ($\pm \text{nan}$)	0.778 (± 0.066)
		VI: Structured Normal	0.812 ($\pm \text{nan}$)	1.040 (± 0.103)
		VI: IAF	0.802 ($\pm \text{nan}$)	0.846 (± 0.078)
		ICL (ours)	1.686 ($\pm \text{nan}$)	3.477 (± 0.604)
		MAP	0.539 ($\pm \text{nan}$)	0.618 (± 0.054)
		TabPFN	0.322 ($\pm \text{nan}$)	0.534 (± 0.038)

Table 6: 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	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 ($\pm \text{nan}$)	0.521 (± 0.028)
		Laplace Approximation	1.541 ($\pm \text{nan}$)	0.655 (± 0.040)
		VI: DiagonalNormal	1.576 ($\pm \text{nan}$)	0.639 (± 0.041)
		VI: MultivariateNormal	1.659 ($\pm \text{nan}$)	0.592 (± 0.035)
		VI: Structured Normal	2.076 ($\pm \text{nan}$)	1.018 (± 0.102)
		VI: IAF	1.706 ($\pm \text{nan}$)	0.627 (± 0.040)
		ICL (ours)	10.319 ($\pm \text{nan}$)	1.458 (± 0.193)
		MAP	1.318 ($\pm \text{nan}$)	0.416 (± 0.018)
		TabPFN	0.330 ($\pm \text{nan}$)	0.443 (± 0.024)