Supplementary Materials for "Classifier Design Given an Uncertainty Class of Feature Distributions via Regularized Maximum Likelihood and the Incorporation of Biological Pathway Knowledge in Steady-State Phenotype Classification"

Mohammad Shahrokh Esfahani[a], Jason Knight[a], Amin Zollanvari[a], Byung-Jun Yoon[a], Edward R. Dougherty[a,b]

[a] *Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX USA*
[b] *The Computational Biology Division of the Translational Genomics Research Institute, Phoenix, AZ USA*

---

1. Variance of uncertainty classes

1.1. $\gamma-$contamination class

In the following, we use γ instead of ε in the main paper.

Using equation (28) in the paper, and knowing that each state's steady-state probability has a beta distribution, Beta$(1, b-1)$, we obtain

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\pi}}^y_{\mathbf{uc}}) &= \sum_{k=1}^{b} \tfrac{1}{|\Pi^y|} \sum_{i=1}^{|\Pi^y|} \mathrm{Var}(\pi^y_i(k)) \\ &= \sum_{k=1}^{b} \tfrac{1}{|\Pi^y|} \sum_{i=1}^{|\Pi^y|} \Big[(1-\gamma_y)^2 \mathrm{Var}(\pi^y_{ac}(k)) + \gamma_y^2 \mathrm{Var}(\pi_i(k))\Big] \\ &= \sum_{k=1}^{b} \tfrac{1}{|\Pi^y|} \sum_{i=1}^{|\Pi^y|} \gamma_y^2 \tfrac{b-1}{b^2(b+1)} = \tfrac{\gamma_y^2 b(b-1)}{b^2(b+1)}.\end{aligned} \tag{1}$$

1.2. $p-$point class

Using the defined mapping in equation (12), for state $k = 1, ..., b$, we obtain

$$\mathrm{Var}(\hat{\pi}^y_{\mathbf{uc}}(k)) = \frac{(\omega^y_{P^y(k)})^2 (|s^y_{P^y(k)}| - 1)}{|s^y_{P^y(k)}|^2 (|s^y_{P^y(k)}| + 1)}. \tag{2}$$

Since states which belong to the same partition contribute equally to the variance of the whole distribution, we can write

$$\mathrm{Var}(\hat{\boldsymbol{\pi}}^y_{\mathbf{uc}}) = \sum_{p=1}^{m_y} \frac{(\omega^y_p)^2 |s^y_p| (|s^y_p| - 1)}{|s^y_p|^2 (|s^y_p| + 1)}. \tag{3}$$

2. RML performance on $\gamma-$contamination and $p-$point uncertainty classes

In this section, we show the performance of the RML-classifier, $\psi_{\mathbf{RML}}$, assuming different structures for the underlying uncertainty class: $\gamma-$contamination and $p-$point class.

We consider two scenarios:

- Exact heuristic regularization parameters: In this case the variance of the data estimate of the conditional distributions, for $y \in \{0, 1\}$, is given by

$$\mathrm{Var}(\hat{\boldsymbol{\pi}}^y_{\mathbf{data}}) = \sum_{k=1}^{b} \frac{\pi^y_{ac}(k)(1 - \pi^y_{ac}(k))}{n_y}. \tag{4}$$

  The variance of the uncertainty classes are given in equations (1) and (3), for $\gamma-$contamination and $p-$point uncertainty classes, respectively.

- Estimated heuristic regularization parameters: In this case, we estimate the entities in equations (1), (3), and (4) using the given sample and uncertainty classes as follows

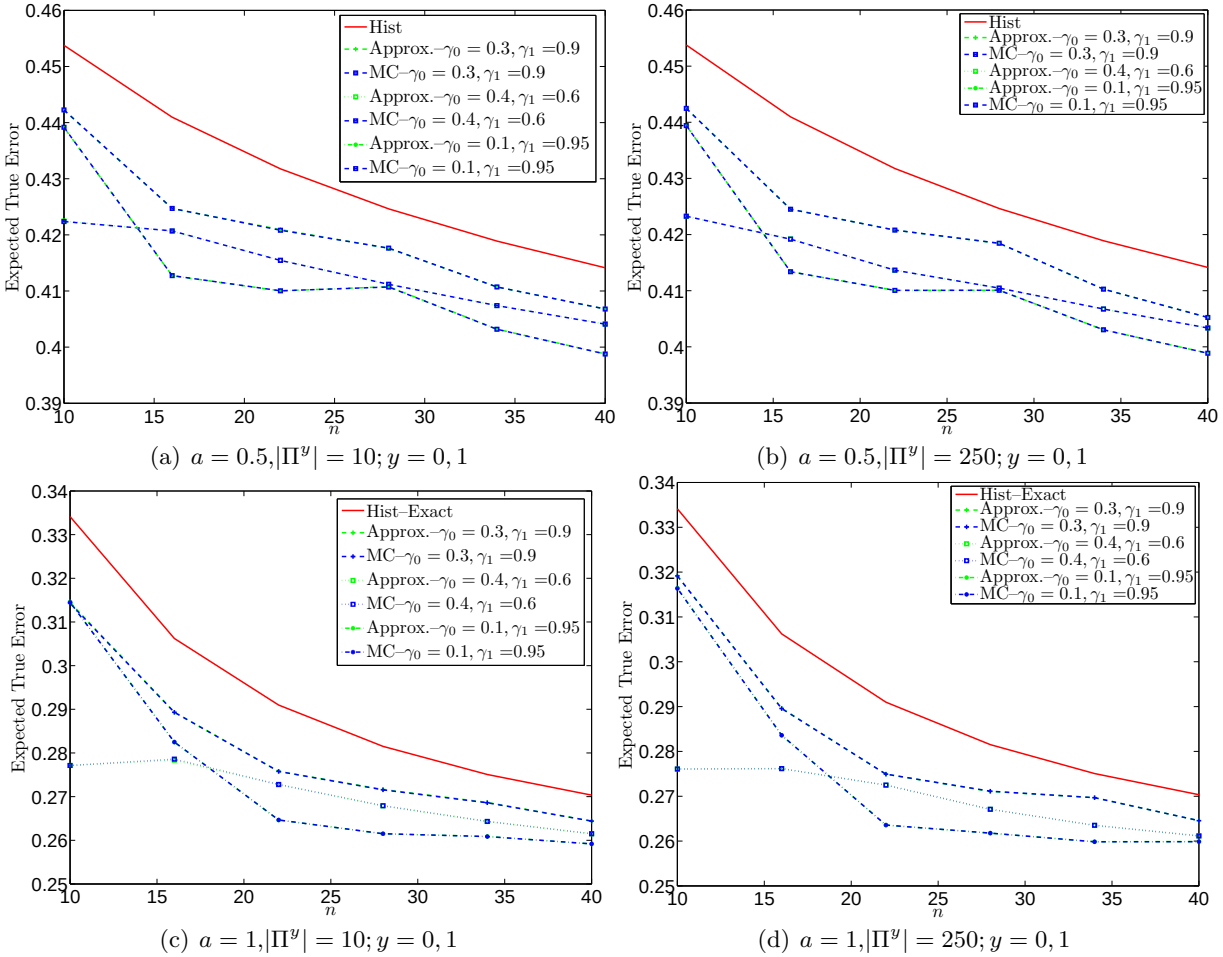
$$\hat{\mathrm{Var}}(\hat{\boldsymbol{\pi}}^y_{\mathbf{uc}}) := \frac{1}{|\Pi^y| - 1} \sum_{k=1}^{b} \sum_{i=1}^{|\Pi^y|} (\pi^y_i(k) - \hat{\boldsymbol{\pi}}^y(k))^2, \tag{5}$$

$$\hat{\mathrm{Var}}(\hat{\boldsymbol{\pi}}^y_{\mathbf{data}}) := \sum_{k=1}^{b} \frac{\frac{u^y_k}{n_y}(1 - \frac{u^y_k}{n_y})}{n_y}. \tag{6}$$

2.1. Results for $\gamma-$contamination uncertainty class

Figures 1(a)-1(d) show the results for the first scenario. Three cases are considered for the pair (γ_0, γ_1): $(0.3, 0.9)$, $(0.4, 0.6)$, and $(0.1, 0.95)$. The expected true error of the RML scheme is smaller than that of the histogram rule in all cases, which shows that the RML paradigm outperforms the traditional histogram rule. Moreover, the results from the Monte-Carlo simulations are very close to those obtained from our approximations in Theorems 1-3 shown by "Approx." in the legends of plots.

Figure 1 shows that the expected true error for the case $(0.4, 0.6)$ is significantly smaller than the others for small sample sizes. This is due to the reliable prior knowledge compared to other cases, for small samples. However, when the sample size increases, $(0.1, 0.95)$ outperforms $(0.4, 0.6)$. Owing to a small contamination degree γ_0 in $(0.1, 0.95)$, the RML framework provides a pretty accurate estimate of $\boldsymbol{\pi}^0_{ac}$ for any sample size. Furthermore, by increasing the sample size, we achieve a better



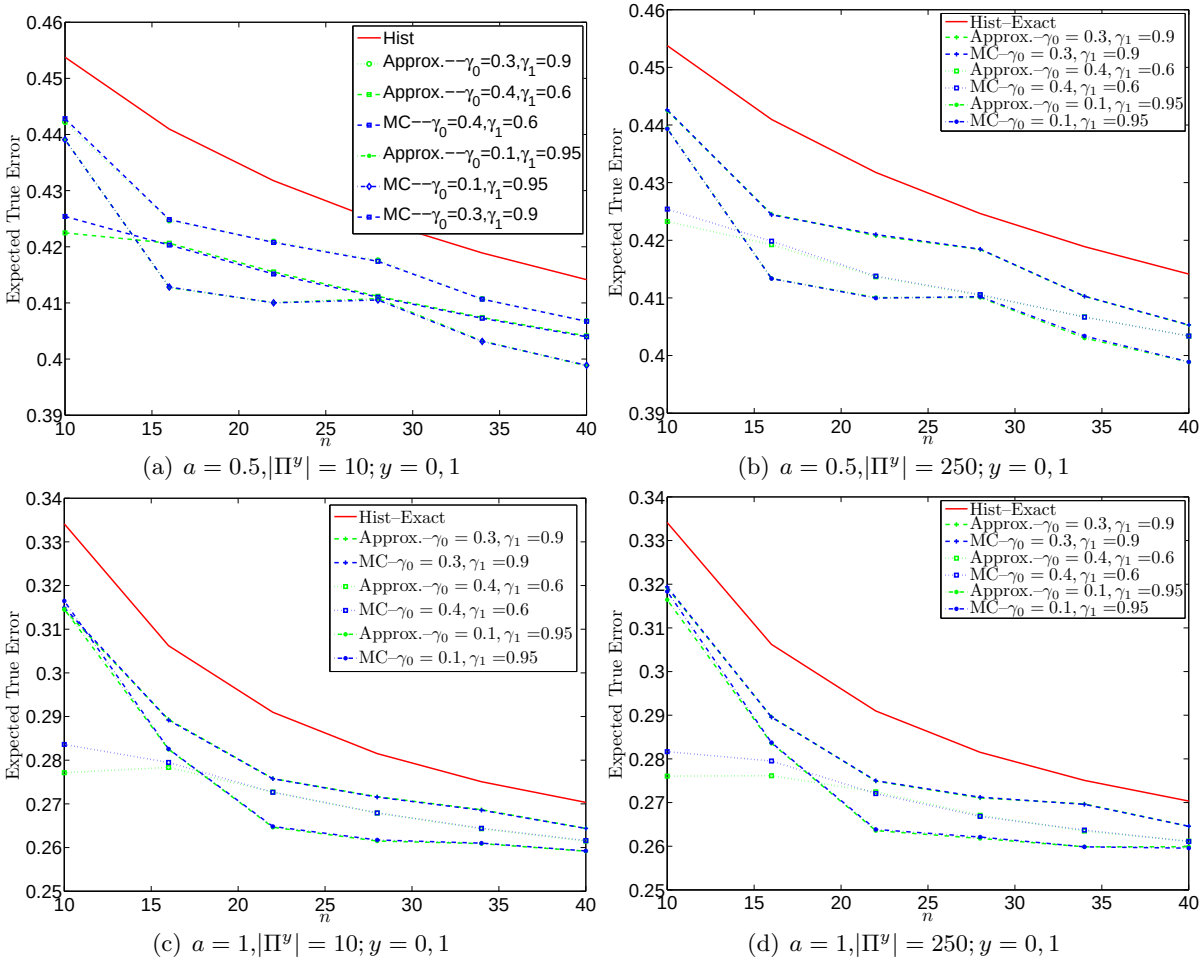
(a) $a = 0.5, |\Pi^y| = 10; y = 0, 1$

(b) $a = 0.5, |\Pi^y| = 250; y = 0, 1$

(c) $a = 1, |\Pi^y| = 10; y = 0, 1$

(d) $a = 1, |\Pi^y| = 250; y = 0, 1$

Figure 1: Expected true error of the RML classifier and Histogram rule with $\gamma-$contamination uncertainty classes. Steady-state distributions with $b = 2^3$ states are considered. The horizontal axis shows the sample size n. Three values for the pair (γ_0, γ_1) are considered. In (a)-(d) the regularization parameters, λ_y, $y = 0, 1$, are exact and computed using equations (1) and (4).



(a) $a = 0.5, |\Pi^y| = 10; y = 0, 1$

(b) $a = 0.5, |\Pi^y| = 250; y = 0, 1$

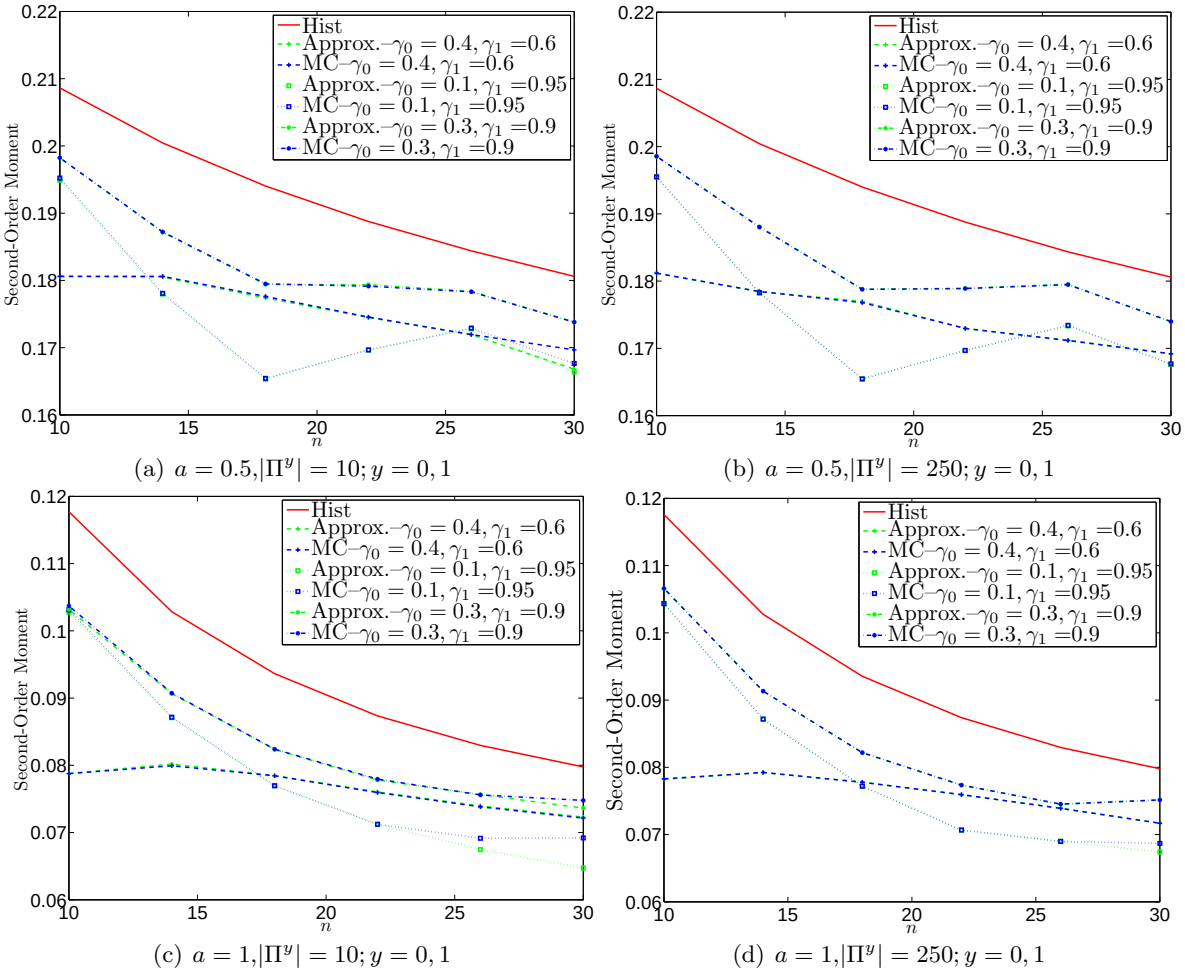
(c) $a = 1, |\Pi^y| = 10; y = 0, 1$

(d) $a = 1, |\Pi^y| = 250; y = 0, 1$

Figure 2: Expected true error of the RML classifier and Histogram rule with $\gamma-$contamination uncertainty classes. Steady-state distributions with $b = 2^3$ states are considered. The horizontal axis shows the sample size n. Three values for the pair (γ_0, γ_1) are considered. In (a)-(d) the regularization parameters, λ_y, $y = 0, 1$, are estimated using equations (5) and (6).

estimate of $\boldsymbol{\pi}^1_{ac}$, making the designed classifier perform close to the optimal classifier. Therefore, it outperforms $(0.4, 0.6)$, which has less accurate estimates of the two conditional distributions for these sample sizes.

Figures 2(a)-2(d) show the results for the data-dependent regularization parameter, where one can see that our approximation and the Monte-Carlo simulations are slightly different. This happens only for small sample sizes, owing to having a poor estimate of $\lambda_y; y = 0, 1$ defined in (4)-(**??**).



(a) $a = 0.5, |\Pi^y| = 10; y = 0, 1$

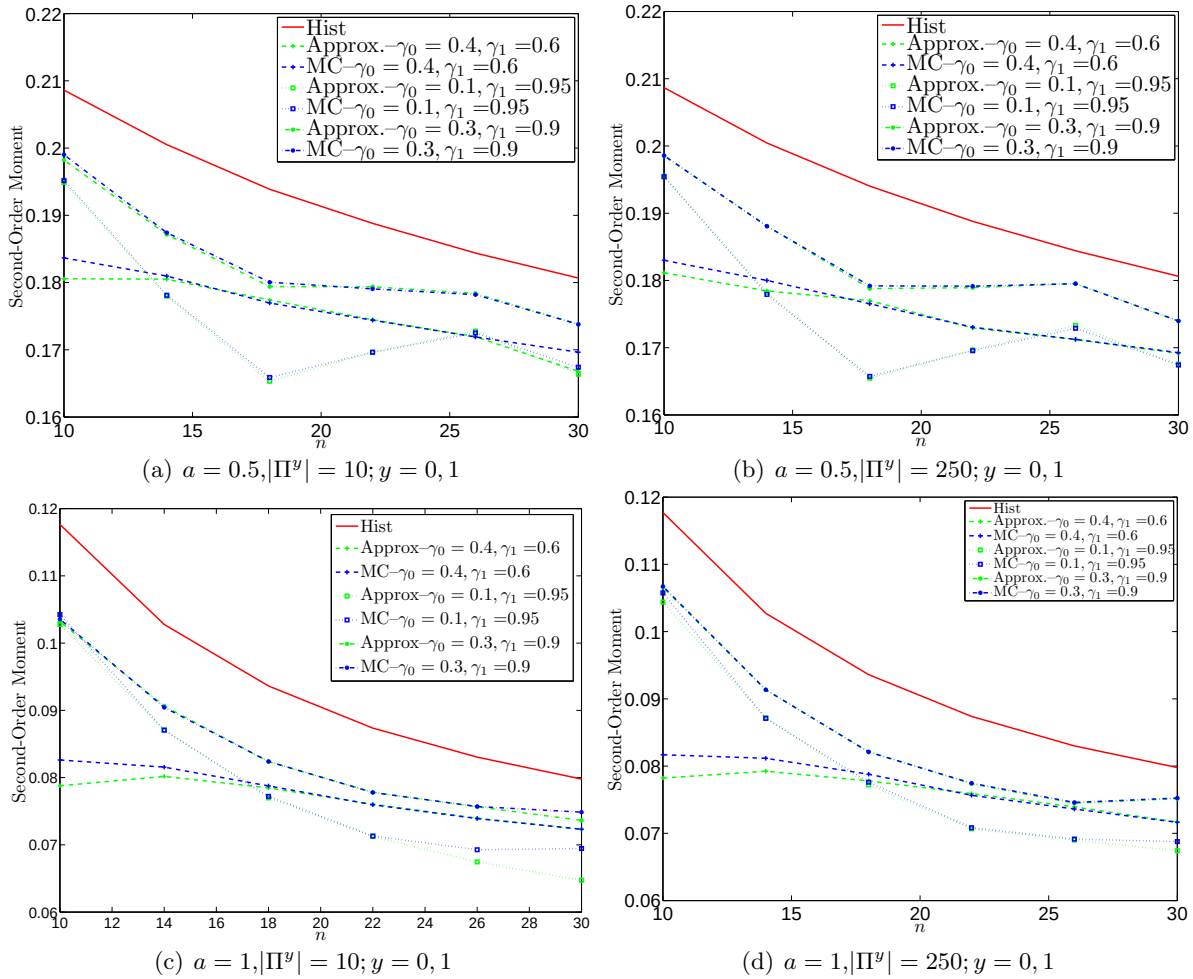
(b) $a = 0.5, |\Pi^y| = 250; y = 0, 1$

(c) $a = 1, |\Pi^y| = 10; y = 0, 1$

(d) $a = 1, |\Pi^y| = 250; y = 0, 1$

Figure 3: Second-order moment of the true error of the RML classifier with $\gamma-$contamination uncertainty class. Steady-state distributions with $b = 2^3$ states are considered. The horizontal axis shows the sample size n. Three values for the pair (γ_0, γ_1) are considered. In (a)-(d) the regularization parameters, λ_y, $y = 0, 1$, are exact and computed using equations (1) and (4).

Similarly, for the second-order moment, we considered two aforementioned scenarios where



(a) $a = 0.5, |\Pi^y| = 10; y = 0, 1$

(b) $a = 0.5, |\Pi^y| = 250; y = 0, 1$

(c) $a = 1, |\Pi^y| = 10; y = 0, 1$

(d) $a = 1, |\Pi^y| = 250; y = 0, 1$

Figure 4: Second-order moment of the true error of the RML classifier with $\gamma-$contamination uncertainty class. Steady-state distributions with $b = 2^3$ states are considered. The horizontal axis shows the sample size n. Three values for the pair (γ_0, γ_1) are considered. In (a)-(d) the regularization parameters, λ_y, $y = 0, 1$, are estimated using equations (5) and (6).

Table 1: Different partitioning with $b = 8$ used in generating $p-$point uncertainty classes.

Partition	Zipf model parameter (a)	$\mathrm{Var}(\boldsymbol{\pi}^0_{\mathbf{uc}})$	$\mathrm{Var}(\boldsymbol{\pi}^1_{\mathbf{uc}})$
$\mathcal{P}^0_1 = \{\{1,2,3,4\},\{5,6,7,8\}\}$			
$\mathcal{P}^1_1 = \{\{1,3,5,7\},\{2,4,6,8\}\}$	0.5	0.0806	0.0757
$\mathcal{P}^0_2 = \{\{1\},\{2,3,4,5,7\},\{6,8\}\}$			
$\mathcal{P}^1_2 = \{\{1,2,3\},\{4,5,6,7,8\}\}$	0.5	0.0526	0.0842
$\mathcal{P}^0_1 = \{\{1,2,3,4\},\{5,6,7,8\}\}$			
$\mathcal{P}^1_1 = \{\{1,3,5,7\},\{2,4,6,8\}\}$	1	0.0963	0.0791
$\mathcal{P}^0_2 = \{\{1\},\{2,3,4,5,7\},\{6,8\}\}$			
$\mathcal{P}^1_2 = \{\{1,2,3\},\{4,5,6,7,8\}\}$	1	0.0386	0.0984

results are shown in Figures 3- 4.

2.1.1. Results for $p-$point class

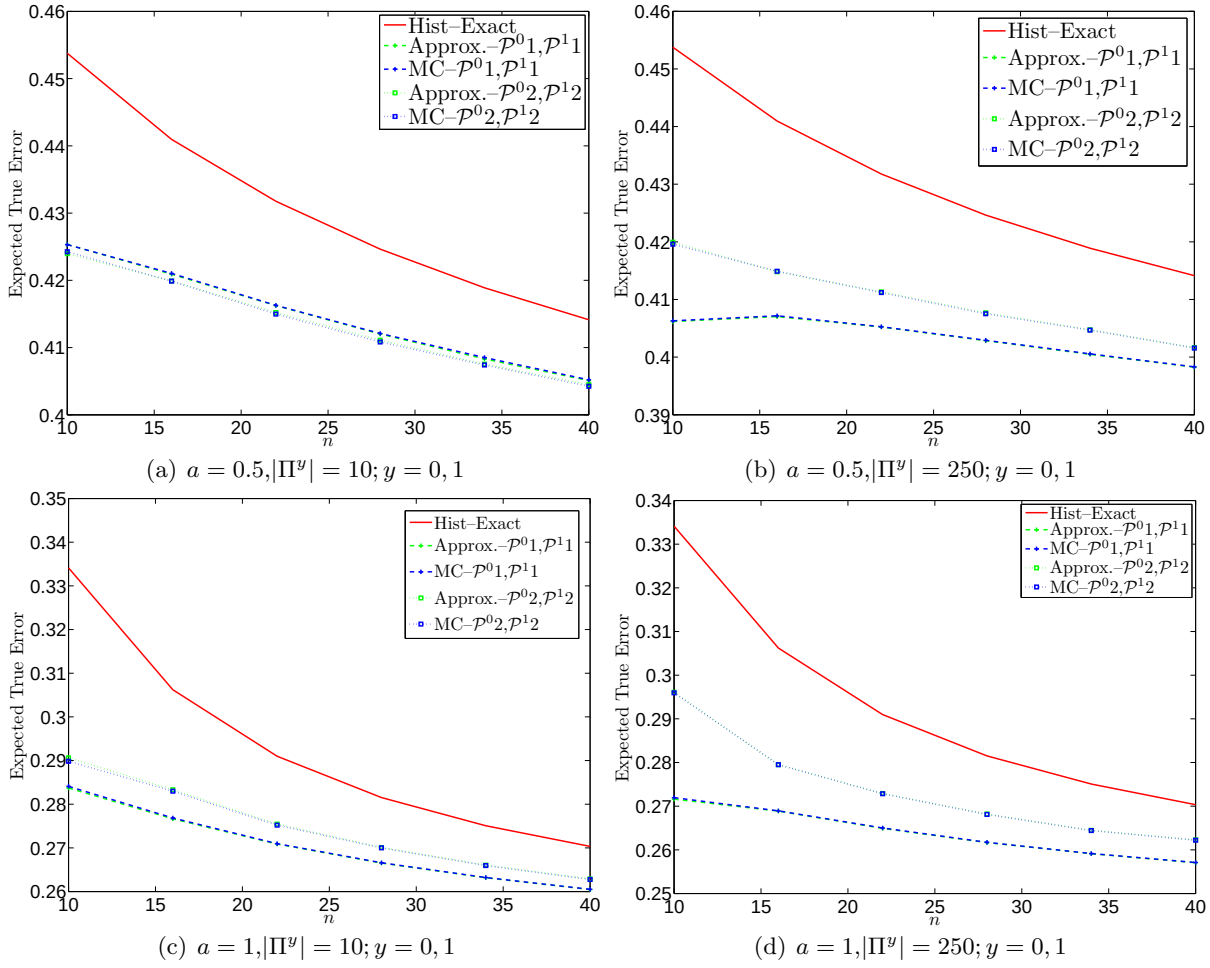
The expected true error of the $p-$point uncertainty classes are illustrated in Figures 5- 6, respectively. As described previously, we use the algorithm proposed in [1] for generating the steady-state distribution corresponding to each partition shown in Table 1. We have considered the two different partitioning cases shown in Table 1. In Table 1, we illustrate the corresponding variance which depends on the partition and the Zipf-model parameter.

Assuming the $p-$point uncertainty class for the underlying uncertainty classes, one can see that the RML classifier outperforms the histogram-classifier for all sample sizes shown in Figure 5. Moreover, similar to the $\gamma-$contamination class, our approximate performs very well for the first scenario (i.e., when we employ the exact regularization parameter obtained by our heuristic approach in subsection 4.3). However, estimating the regularization parameter happens to slightly degrade the performance of the approximation.

Similarly, for the second-order moment, we considered two aforementioned scenarios where results are shown in Figures 7- 8.

3. Generating uncertainty classes from pathways

We provide a simple example of how a set of biological pathways can generate an uncertainty class of stochastic network models. Consider three pathways describing the dynamical behavior of



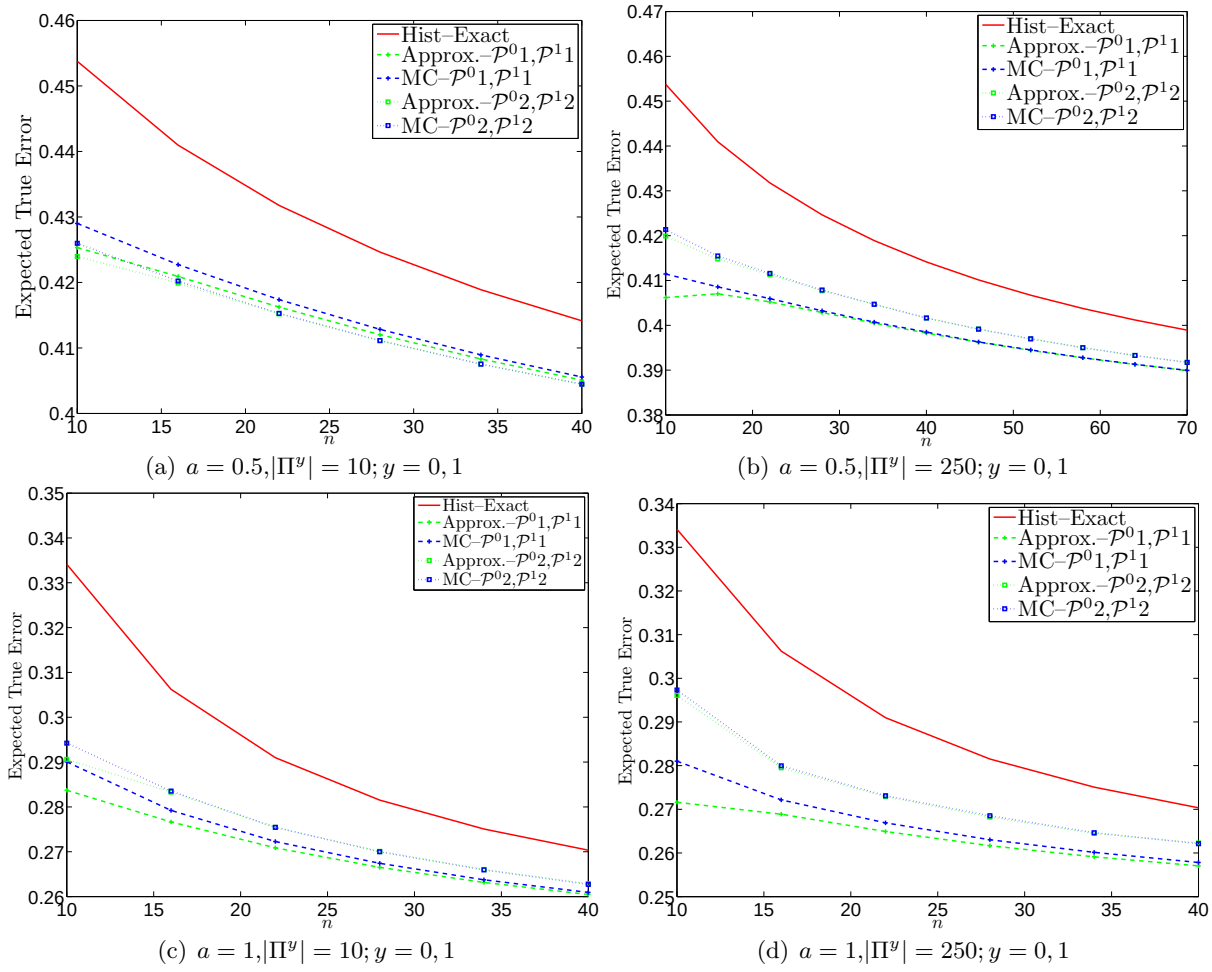
(a) $a = 0.5, |\Pi^y| = 10; y = 0, 1$

(b) $a = 0.5, |\Pi^y| = 250; y = 0, 1$

(c) $a = 1, |\Pi^y| = 10; y = 0, 1$

(d) $a = 1, |\Pi^y| = 250; y = 0, 1$

Figure 5: Expected true error of the RML and Histogram with the $p-$point uncertainty classes described in Table 1. Steady-state distributions with $b = 2^3$ states are considered. The horizontal axis shows the sample size n. In (a)-(d) the regularization parameters, λ_y, $y = 0, 1$, are exact and computed using equations (3) and (4).



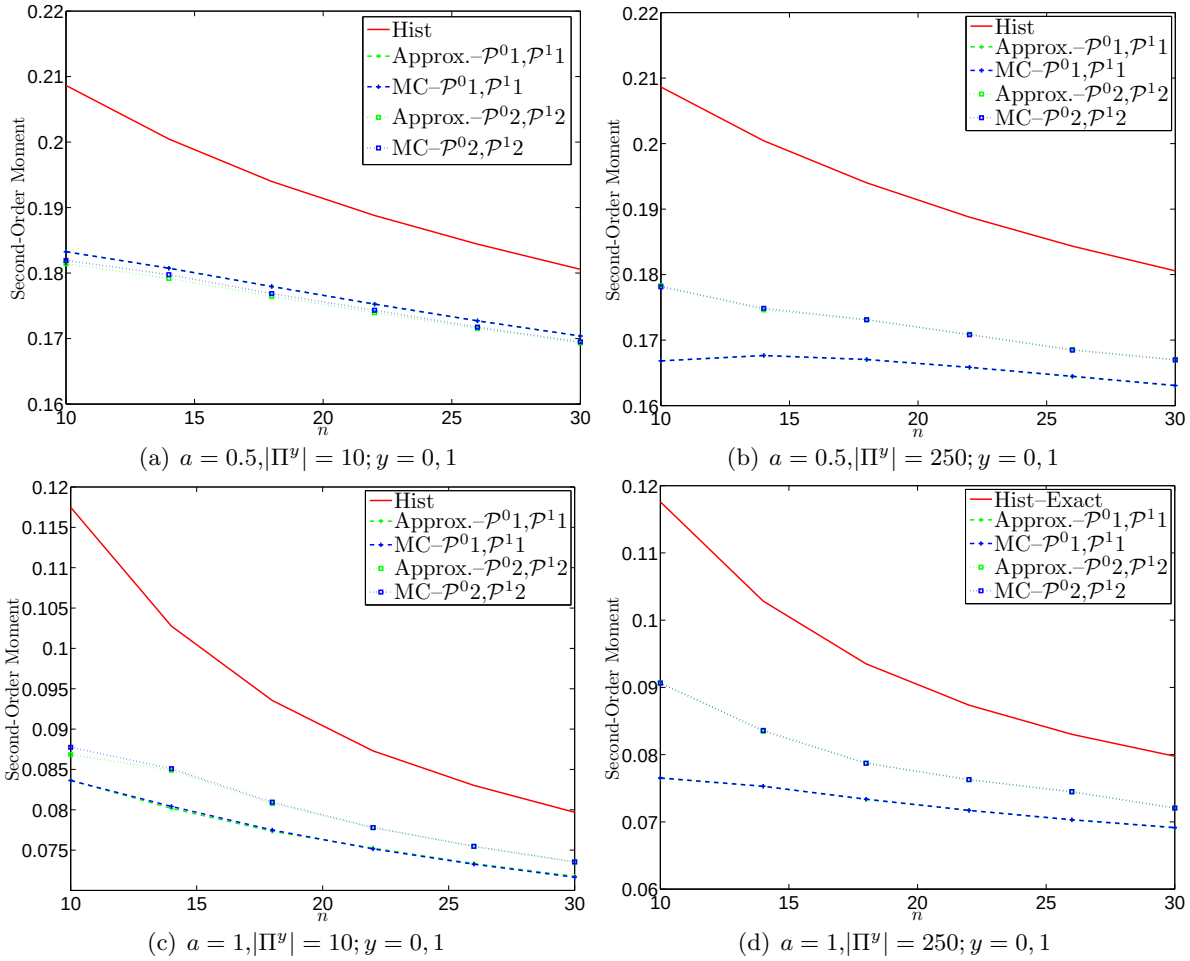
(a) $a = 0.5, |\Pi^y| = 10; y = 0, 1$

(b) $a = 0.5, |\Pi^y| = 250; y = 0, 1$

(c) $a = 1, |\Pi^y| = 10; y = 0, 1$

(d) $a = 1, |\Pi^y| = 250; y = 0, 1$

Figure 6: Expected true error of the RML and Histogram with the $p-$point uncertainty classes described in Table 1. Steady-state distributions with $b = 2^3$ states are considered. The horizontal axis shows the sample size n. In (a)-(d) the regularization parameters, λ_y, $y = 0, 1$, are estimated using equations (5) and (6).



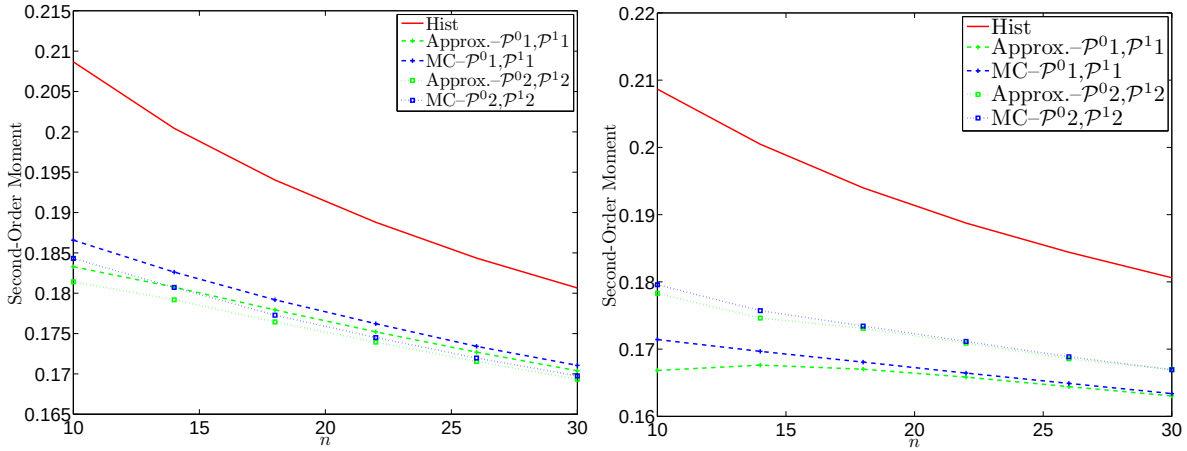
(a) $a = 0.5, |\Pi^y| = 10; y = 0, 1$

(b) $a = 0.5, |\Pi^y| = 250; y = 0, 1$

(c) $a = 1, |\Pi^y| = 10; y = 0, 1$

(d) $a = 1, |\Pi^y| = 250; y = 0, 1$

Figure 7: Second-order moment of the true error of the RML classifier with $p-$point uncertainty class described in Table 1. Steady-state distributions with $b = 2^3$ states are considered. The horizontal axis shows the sample size n. In (a)-(d) the regularization parameters, λ_y, $y = 0, 1$, are exact and computed using equations (3) and (4).



(a) $a = 0.5, |\Pi^y| = 10; y = 0, 1$

(b) $a = 0.5, |\Pi^y| = 250; y = 0, 1$

(c) $a = 1, |\Pi^y| = 10; y = 0, 1$

(d) $a = 1, |\Pi^y| = 250; y = 0, 1$

Figure 8: Second-order moment of the true error of the RML classifier with $p-$point uncertainty class described in Table 1. Steady-state distributions with $b = 2^3$ states are considered. The horizontal axis shows the sample size n. In (a)-(d) the regularization parameters, λ_y, $y = 0, 1$, are estimated using equations (5) and (6).

two binary genes A and B:

$$B = 1 \implies A = 0 \tag{7}$$

$$B = 0 \implies A = 1 \tag{8}$$

$$A = 1 \implies B = 1 \tag{9}$$

This simple system is almost completely specified, but when gene A is in state 0, the dynamical behavior of gene B is unspecified. In [2], Layek *et al.* show how to generate an uncertainty class of deterministic networks from a set of pathways by relaxing timing considerations. Knight *et al.* in [3] use a stochastic approach to generate a single Markov chain from a set of pathways and validate the approach using pathways for the NF-κB transcription factor system. In [4] it is shown that the earlier approach in [3] can be generalized to produce a parameterized uncertainty class of Markov chains from a given set of pathways.

Based on [4] we generate the parameterized state transition graph in Fig. 9. By choosing $\theta \in [0, 1]$ we fix the stochastic evolution of gene B and this characterizes a single Markov chain. We can therefore think of the graph in Fig. 9 as the state transition graph of an entire uncertainty class of Markov chains. The regularized maximum likelihood classification technique requires a

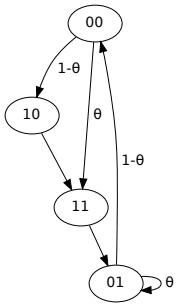
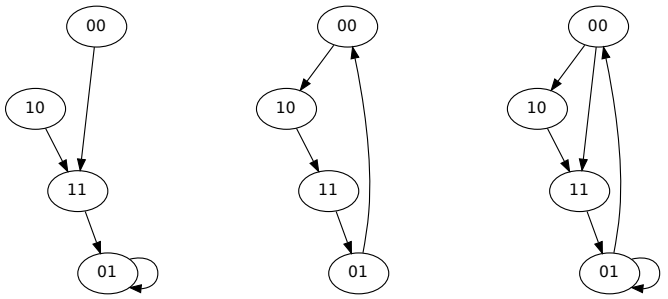


Figure 9: The parameterized state transition graph of a two gene, four state Markov chain system derived from three pathways. The node labels should be read [A,B]. The parameter θ determines the evolution of gene B when gene A= 0.

finite uncertainty class so we effectively sample this uncountably infinite, parameterized uncertainty class of Markov chains by discretizing the values of θ. Specifically in this simple example, we use $\theta \in \{0, 0.5, 1\}$ to consider three networks where the behavior of gene B is deterministically up-

regulated, deterministically down-regulated, and is a mixture of the two behaviors. The mixture case can be understood to encompass more complex biological regulation such as time-varying pulses or more complex, nonlinear stochastic regulation where we only care about the long-run activity. These three networks are shown in Figure 10.



(a) (b) (c)

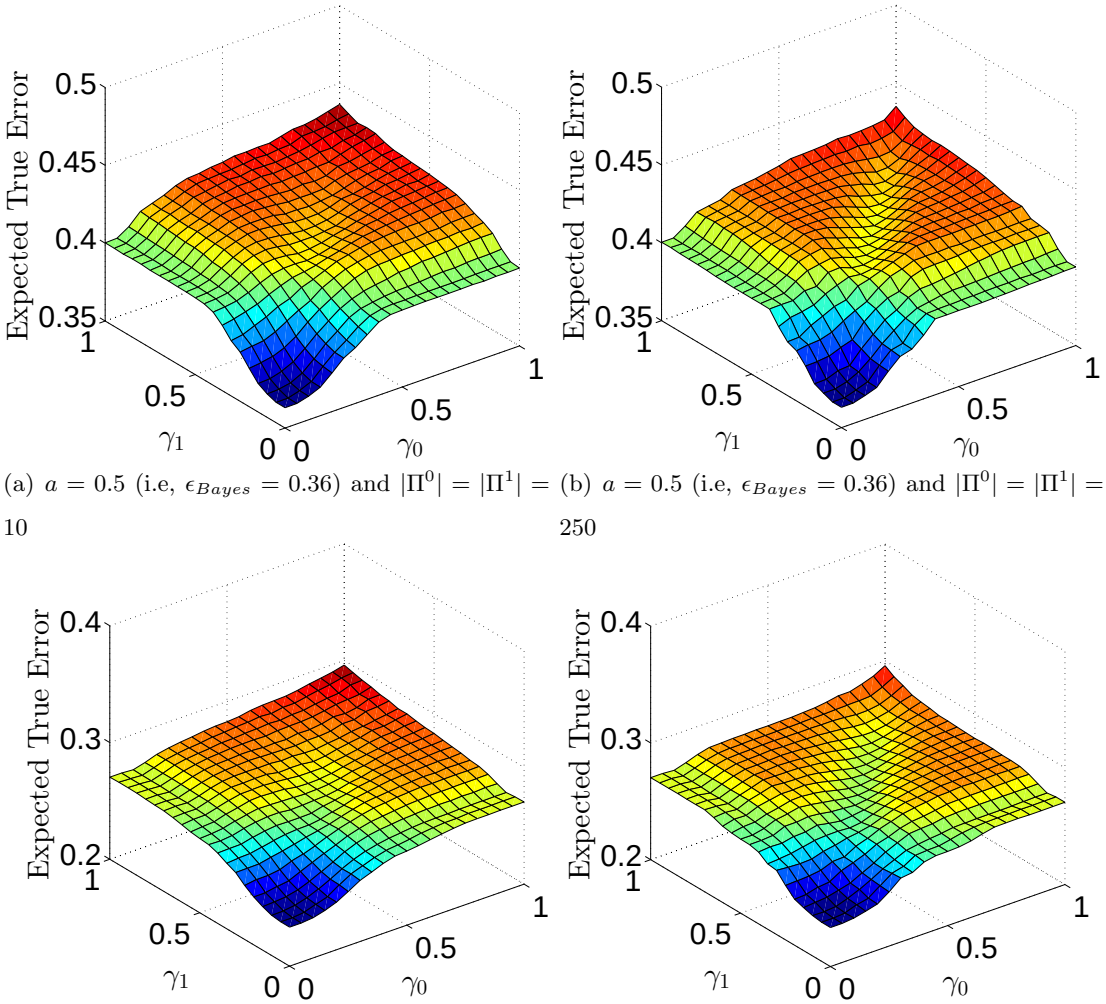
Figure 10: The state transition graphs of the Markov chains when (10(a)) $\theta = 0$, (10(b)) $\theta = 1$, and (10(c)) $\theta = 0.5$ where all outgoing edges from a given node are equiprobable. Depending on the value of θ the network can have a singleton atttractor state, a large attractor cycle, or a mixture of these two long run behaviors.

4. Effect of amount of contamination on the expected true error of the RML classifier

In this section, we observe the effect of the uncertainty, e.g., contamination in a $\gamma-$contamination uncertainty class, on the performance of the RML-classifier. The following figures show how the amount of uncertainty affects the expected true error of the RML classifier. In this study, we considered $\gamma-$contamination model for the uncertainty classes. γ_0 and γ_1 determine degree of contamination for Π^0 and Π^1, respectively. Figures 11-12 show the expected true error surface as a function of two contamination degrees, γ_0 and γ_1. For these figures, we grid the rectangle $[0,1]^2$ into 400 pairs of (γ_0, γ_1) and computed the expected true error which can be found in Theorem 1 of the paper. Figure 11 shows the results for $n = 20 : n_0 = n_1 = 10$. Results for $n = 40 : n_0 = n_1 = 20$ are shown in Figure 12.

One can see that in all plots, when we have $(\gamma_0, \gamma_1) = (0, 0)$, we reach the Bayes error. This is due to the fact that the uncertainty classes only contain the true label-conditional distributions and therefore, the regularization parameters, λ_0 and λ_1, will be one. It means that, we do not use

training data and only use our prior knowledge about the SSD. On the other hand, increasing the contamination degree to the corner point $(\gamma_0, \gamma_1) = (1, 1)$, degrades the performance. However, one can see that this is not significant, because the RML framework takes the amount of uncertainty into account by tuning the regularization parameter accordingly.



(a) $a = 0.5$ (i.e, $\epsilon_{Bayes} = 0.36$) and $|\Pi^0| = |\Pi^1| = 10$

(b) $a = 0.5$ (i.e, $\epsilon_{Bayes} = 0.36$) and $|\Pi^0| = |\Pi^1| = 250$

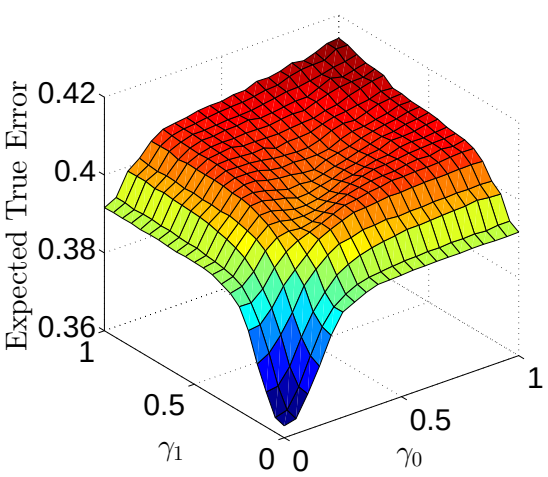
(c) $a = 1$ (i.e, $\epsilon_{Bayes} = 0.23$) and $|\Pi^0| = |\Pi^1| = 10$

(d) $a = 1$ (i.e, $\epsilon_{Bayes} = 0.23$) and $|\Pi^0| = |\Pi^1| = 250$

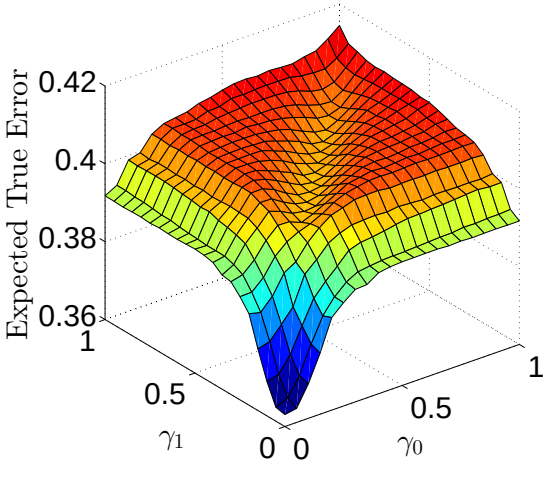
Figure 11: Expected true error of the RML with $n_0 = n_1 = 10$ sample points. We considered $b = 8$ bins..

References

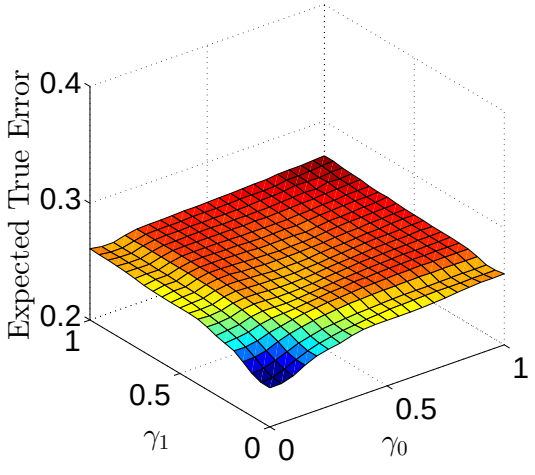
[1] N. Smith and R. Tromble, "Sampling uniformly from the unit simplex," *Johns Hopkins University, Tech. Rep*, 2004.



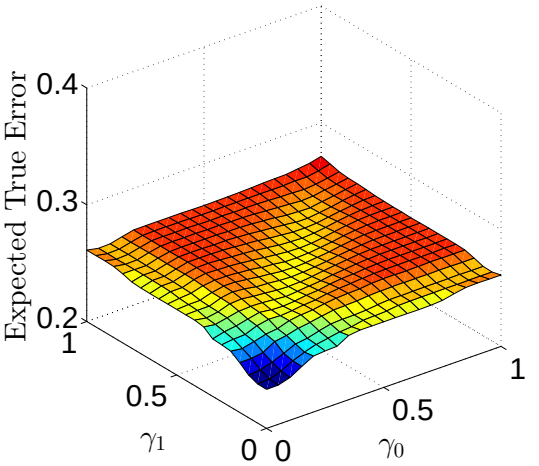
(a) $a = 0.5$ (i.e, $\epsilon_{Bayes} = 0.36$) and $|\Pi^0| = |\Pi^1| = 10$



(b) $a = 0.5$ (i.e, $\epsilon_{Bayes} = 0.36$) and $|\Pi^0| = |\Pi^1| = 250$



(c) $a = 0.5$ (i.e, $\epsilon_{Bayes} = 0.23$) and $|\Pi^0| = |\Pi^1| = 10$



(d) $a = 0.5$ (i.e, $\epsilon_{Bayes} = 0.23$) and $|\Pi^0| = |\Pi^1| = 250$

Figure 12: Expected true error of the RML with $n_0 = n_1 = 20$ sample points. We considered $b = 8$ bins.

[2] R. Layek, A. Datta, and E. Dougherty, "From biological pathways to regulatory networks," *Mol. BioSyst.*, vol. 7, no. 3, pp. 843–851, 2011.

[3] J. Knight, A. Datta, and E. Dougherty, "A stochastic nf-κb model consistent with pathway information," *In Press, Transaction on Biomedical Engineering, IEEE*, 2012.

[4] J. Knight and E. Dougherty, "Attractor estimation and model refinement for stochastic regulatory network models," *Genomic Signal Processing and Statistics (GENSIPS)*, pp. 54–55, 2011.