Universal Outlier Hypothesis Testing

Sirin Nitinawarat

ECE Dept & CSL

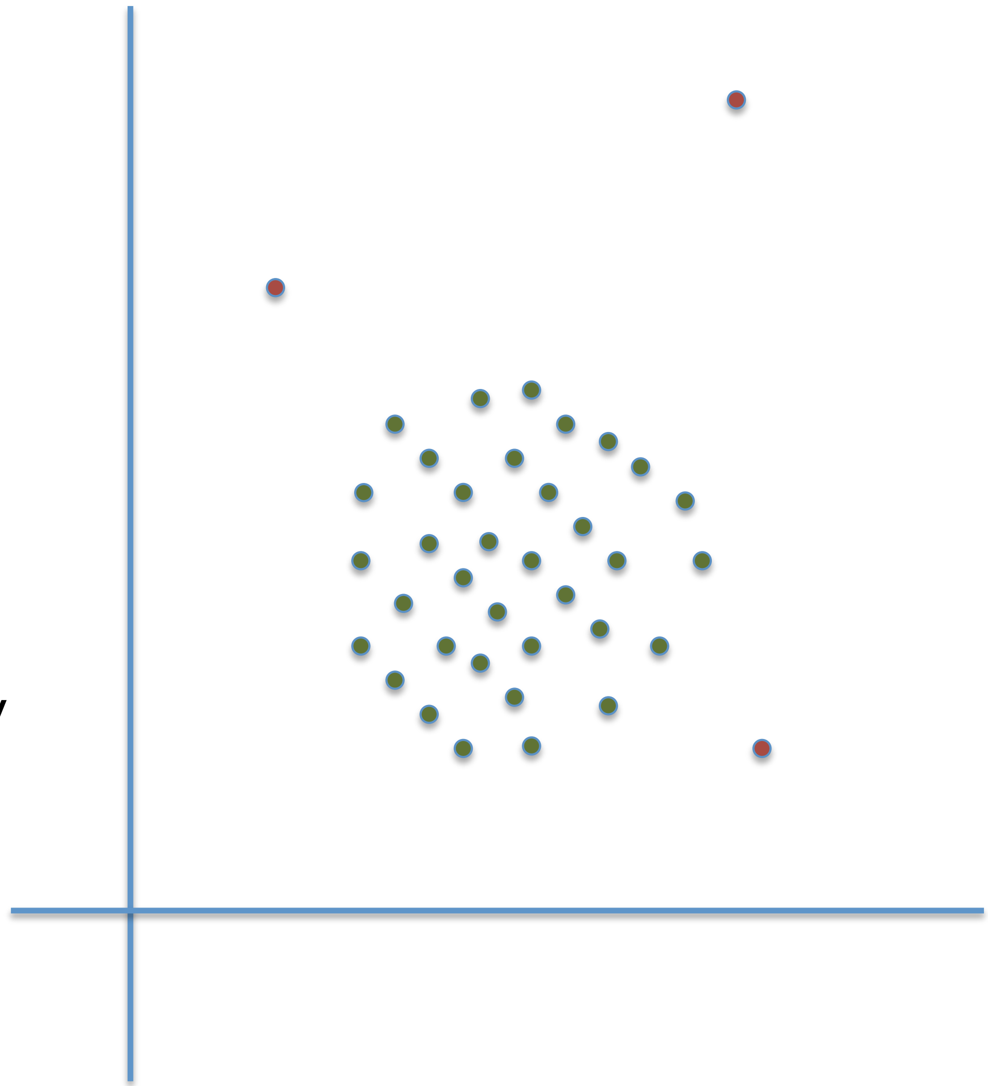
University of Illinois at Urbana-Champaign

http://www.ifp.illinois.edu/~nitinawa

(with Yun Li and Prof. Venu V. Veeravalli)

Statistical Outlier Detection

- Single sequence of observations
- Generic observations follow some fixed (possibly unknown) distribution or generating mechanism
- Outliers follow different generating mechanism
- Goal: To find outliers efficiently
- Applications: fraud detection, public health monitoring, cleaning up data



Fraud Detection

- Example: spending records for a male graduate student

Trans-actions	Grocery	Gas	---	Books	Grocery		
Amount	30$	35$	---	75$	35$		

Generic behavior

Fraud Detection

- Example: spending records for a male graduate student

Trans-actions	Grocery	Gas	---	Books	Grocery	Spa	Cosmetics
Amount	30$	35$	---	75$	35$	250$	500$

Generic behavior

Fraudulent behavior

Fraud Detection: Group Monitoring

- Male graduate students

Student 1	Student 2	Student 3	…	Student *M*
Grocery	Dining	Grocery	…	Gas
Dining	Grocery	Gas	…	Books
…	…	…	…	…
Books	Movie	Books	…	Grocery
Movie	Books	Grocery	…	Dining
Grocery	Gas	spa	…	Movie
Gas	Books	Cosmetics	…	Grocery

Outlier Hypothesis Testing

- *M* sequences of observations, with *M* large
- Almost all sequences are generated from common typical distribution
- Small subset of sequences generated from different (outlier) distribution

Outlier Hypothesis Testing

- M sequences of observations, with M large
- Almost all sequences are generated from common typical distribution
- Small subset of sequences generated from different (outlier) distribution
- Special case:
  - Exactly one sequence is generated from outlier distribution
  - Goal: to detect outlier sequence efficiently
  - Universal setting: neither typical nor outlier distributions known; no training data provided

Universal Outlier Hypothesis Testing

- Typical distribution π
- Outlier distribution μ

	1	2			M
H_1	μ	π	π	π	π
H_2	π	μ	π	π	π
	π	π	μ	π	π
	π	π	π	μ	π
H_M	π	π	π	π	μ

Applications: Outlier Hypothesis Testing

- Search problems and target tracking
- Sensor network applications: event detection, environment monitoring
- Fraud detection and anomaly detection in big data

Mathematical Model

$$\begin{array}{cccc} \boxed{\begin{matrix} y_1^{(1)} \\ y_2^{(1)} \\ \vdots \\ y_n^{(1)} \end{matrix}} & \boxed{\begin{matrix} y_1^{(2)} \\ y_2^{(2)} \\ \vdots \\ y_n^{(2)} \end{matrix}} & \begin{matrix} \cdots \\ \cdots \\ \\ \cdots \end{matrix} & \boxed{\begin{matrix} y_1^{(M)} \\ y_2^{(M)} \\ \vdots \\ y_n^{(M)} \end{matrix}} \\ \mathbf{y}^{(1)} & \mathbf{y}^{(2)} & & \mathbf{y}^{(M)} \end{array}$$

$$H_i: \quad p_i\left(\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}\right) = \prod_{k=1}^{n} \left[\mu\left(y_k^{(i)}\right) \prod_{j \neq i} \pi\left(y_k^{(j)}\right) \right]$$

Universal Outlier Hypothesis Testing

$$H_i : \quad p_i\left(y^{Mn}\right) = \prod_{k=1}^{n} \left[\mu\left(y_k^{(i)}\right) \prod_{j \neq i} \pi\left(y_k^{(j)}\right) \right]$$

Nothing is known about (μ, π) except that they are distinct

Universal Test: $\quad \delta : \mathcal{Y}^{Mn} \rightarrow \{1, \ldots, M\}$

Universal Outlier Hypothesis Testing

$$H_i : \quad p_i\left(y^{Mn}\right) = \prod_{k=1}^{n} \left[\mu\left(y_k^{(i)}\right) \prod_{j \neq i} \pi\left(y_k^{(j)}\right) \right]$$

Nothing is known about (μ,π) except that they are distinct

Universal Test: $\delta : \mathcal{Y}^{Mn} \rightarrow \{1, \ldots, M\}$

Independent of (μ,π)

Performance Metrics

- Maximal error probability:

$$e\left(\delta, \left(\mu, \pi\right)\right) = \max_{i} \mathbb{P}_i \left\{\delta(\mathbf{y}^{Mn}) \neq i\right\}$$

- Exponent for maximal error probability:

$$\alpha\left(\delta, \left(\mu, \pi\right)\right) = \lim_{n \to \infty} -\frac{1}{n} \log e\left(\delta, \left(\mu, \pi\right)\right)$$

Performance Metrics

- Maximal error probability:

$$e\left(\delta, \left(\mu, \pi\right)\right) = \max_i \mathbb{P}_i\left\{\delta(\mathbf{y}^{Mn}) \neq i\right\}$$

- Exponent for maximal error probability:

$$\alpha\left(\delta, \left(\mu, \pi\right)\right) = \lim_{n\to\infty} -\frac{1}{n}\log e\left(\delta, \left(\mu, \pi\right)\right)$$

Consistency : $e \to 0$ as $n \to \infty$

Exponential Consistency : $\alpha > 0$

Background: Binary Hypothesis Testing

$$H_1 : \; p_1(\mathbf{y}) = \prod_{k=1}^{n} \pi(y_k) \qquad H_2 : \; p_2(\mathbf{y}) = \prod_{k=1}^{n} \mu(y_k)$$

If (μ, π) known, $\delta_{\mathrm{ML}}(\mathbf{y}) = \arg\max_i \log p_i(\mathbf{y})$

has $\alpha(\delta, (\mu, \pi)) = C(\mu, \pi) > 0$

Background: Binary Hypothesis Testing

$$H_1 : \; p_1(\mathbf{y}) \;=\; \prod_{k=1}^{n} \pi(y_k) \qquad H_2 : \; p_2(\mathbf{y}) \;=\; \prod_{k=1}^{n} \mu(y_k)$$

If (μ, π) known, $\delta_{\mathrm{ML}}(\mathbf{y}) \;=\; \arg\max_i \log p_i(\mathbf{y})$

has $\alpha(\delta, (\mu, \pi)) \;=\; C(\mu, \pi) \;>\; 0 \longleftarrow$ exponential consistency

Background: Binary Hypothesis Testing

$$H_1 : \ p_1(\mathbf{y}) = \prod_{k=1}^{n} \pi(y_k) \qquad H_2 : \ p_2(\mathbf{y}) = \prod_{k=1}^{n} \mu(y_k)$$

If (μ, π) known, $\delta_{\mathrm{ML}}(\mathbf{y}) = \arg\max_i \log p_i(\mathbf{y})$

has $\alpha(\delta, (\mu, \pi)) = C(\mu, \pi) > 0 \longleftarrow$ exponential consistency

Chernoff Info : $C(\mu, \pi) = \max_{0 \leq s \leq 1} -\log\left(\sum_y \mu(y)^s \pi(y)^{1-s} \right)$

Outlier Hypothesis Testing: Known (μ, π)

$$H_i: \quad p_i\left(\mathbf{y}^{Mn}\right) = \prod_{k=1}^{n}\left[\mu\left(y_k^{(i)}\right)\prod_{j \neq i}\pi\left(y_k^{(j)}\right)\right]$$

ML Rule : $\quad \delta_{\mathrm{ML}}(y^{Mn}) = \operatorname*{argmax}_i \log p_i(y^{Mn})$

$$\begin{array}{c|ccccc|}
 & 1 & 2 & & & M \\
\hline
H_1 & \mu & \pi & \pi & \pi & \pi \\
H_2 & \pi & \mu & \pi & \pi & \pi \\
 & \pi & \pi & \mu & \pi & \pi \\
 & \pi & \pi & \pi & \mu & \pi \\
H_M & \pi & \pi & \pi & \pi & \mu \\
\hline
\end{array}$$

Exponential Consistency : $\quad \alpha(\delta_{\mathrm{ML}}, (\mu, \pi)) = 2B(\mu, \pi)$

Bhattacharya Distance : $\quad B(\mu, \pi) = -\log\left(\sum_y \mu(y)^{1/2}\pi(y)^{1/2}\right)$

Binary Hypothesis Testing: Unknown μ

$$H_1 : \; p_1(\mathbf{y}) = \prod_{k=1}^{n} \pi(y_k) \qquad H_2 : \; p_2(\mathbf{y}) = \prod_{k=1}^{n} \mu(y_k)$$

If μ unknown

for any given δ there exists μ s.t. $\alpha = 0$

No exponential consistency!

Outlier Hypothesis Testing: Unknown μ

μ unknown : $\hat{\mu}_i = \gamma_i \leftarrow$ empirical distribution

$$H_i : \quad \hat{p}_i\left(y^{Mn}\right) = \prod_{k=1}^{n}\left[\hat{\mu}_i\left(y_k^{(i)}\right)\prod_{j\neq i}\pi\left(y_k^{(j)}\right)\right]$$

Generalized Likelihood (GL) Rule :

$$\delta_{\mathrm{GL}}(y^{Mn}) = \underset{i}{\mathrm{argmax}} \log \hat{p}_i(y^{Mn})$$

$$\begin{array}{c|ccccc|}
 & 1 & 2 & & & M \\
\hline
H_1 & \mu & \pi & \pi & \pi & \pi \\
H_2 & \pi & \mu & \pi & \pi & \pi \\
 & \pi & \pi & \mu & \pi & \pi \\
 & \pi & \pi & \pi & \mu & \pi \\
H_M & \pi & \pi & \pi & \pi & \mu \\
\hline
\end{array}$$

Exponential Consistency : $\alpha(\delta_{\mathrm{GL}}, (\mu, \pi)) = 2B(\mu, \pi)$

Outlier Hypothesis Testing: Unknown μ

μ unknown : $\hat{\mu}_i = \gamma_i \leftarrow$ empirical distribution

$$H_i : \quad \hat{p}_i\left(y^{Mn}\right) = \prod_{k=1}^{n}\left[\hat{\mu}_i\left(y_k^{(i)}\right)\prod_{j \neq i}\pi\left(y_k^{(j)}\right)\right]$$

Generalized Likelihood (GL) Rule :

$$\delta_{\mathrm{GL}}(y^{Mn}) = \underset{i}{\operatorname{argmax}} \log \hat{p}_i(y^{Mn})$$

$$\begin{array}{c|ccccc}
 & 1 & 2 & & & M \\
\hline
H_1 & \mu & \pi & \pi & \pi & \pi \\
H_2 & \pi & \mu & \pi & \pi & \pi \\
 & \pi & \pi & \mu & \pi & \pi \\
 & \pi & \pi & \pi & \mu & \pi \\
H_M & \pi & \pi & \pi & \pi & \mu
\end{array}$$

Exponential Consistency : $\alpha(\delta_{\mathrm{GL}}, (\mu, \pi)) = 2B(\mu, \pi)$

Same as known μ, π

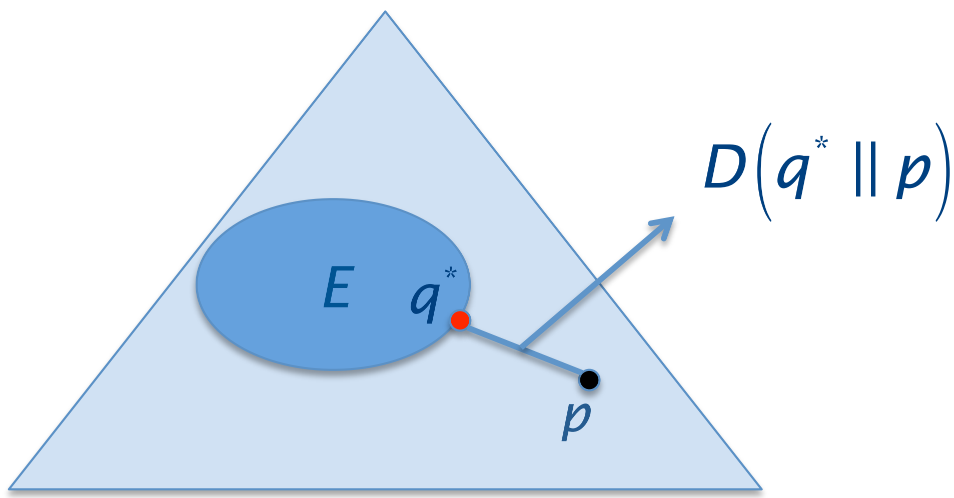
Sanov's Theorem

- Sanov's Theorem: For i.i.d. rvs $Y^n \sim p$, exponent of probability that random empirical distribution falls in closed set E is

Key Tool: Sanov's Theorem

- Sanov's Theorem: For i.i.d. rvs $Y^n \sim p$, exponent of probability that random empirical distribution falls in closed set E is

$$\lim_{n} -\frac{1}{n}\log \mathbb{P}\left\{\text{Empirical}\left(Y^n\right) \in E\right\} = \min_{q\in E} D\left(q \,\|\, p\right)$$



Proposed Universal Test

$$(\mu, \pi) \text{ not known}: \quad \hat{\mu}_i = \gamma_i \qquad \hat{\pi}_i = \frac{1}{M-1} \sum_{j \neq i} \gamma_j$$

empirical distributions

$$H_i: \quad \hat{\hat{p}}_i\left(y^{Mn}\right) = \prod_{k=1}^{n} \left[\hat{\mu}_i\left(y_k^{(i)}\right) \prod_{j \neq i} \hat{\pi}_i\left(y_k^{(j)}\right) \right]$$

Generalized Likelihood (GL) Rule :

$$\delta_{\mathrm{GL}}(y^{Mn}) = \arg\max_i \log \hat{\hat{p}}_i(y^{Mn})$$

$$\begin{array}{c|ccccc|}
 & 1 & 2 & & & M \\
\hline
H_1 & \mu & \pi & \pi & \pi & \pi \\
H_2 & \pi & \mu & \pi & \pi & \pi \\
 & \pi & \pi & \mu & \pi & \pi \\
 & \pi & \pi & \pi & \mu & \pi \\
H_M & \pi & \pi & \pi & \pi & \mu \\
\hline
\end{array}$$

Proposed Universal Test

$$(\mu,\pi) \text{ not known}: \quad \hat{\mu}_i = \gamma_i \qquad \hat{\pi}_i = \frac{1}{M-1}\sum_{j\neq i}\gamma_j$$

$$H_i: \quad \hat{\hat{p}}_i\left(y^{Mn}\right) = \prod_{k=1}^{n}\left[\hat{\mu}_i\left(y_k^{(i)}\right)\prod_{j\neq i}\hat{\pi}_i\left(y_k^{(j)}\right)\right]$$

$$\begin{array}{c|ccccc|}
 & 1 & 2 & & & M \\
\hline
H_1 & \mu & \pi & \pi & \pi & \pi \\
H_2 & \pi & \mu & \pi & \pi & \pi \\
 & \pi & \pi & \mu & \pi & \pi \\
 & \pi & \pi & \pi & \mu & \pi \\
H_M & \pi & \pi & \pi & \pi & \mu \\
\hline
\end{array}$$

Generalized Likelihood (GL) Rule :

$$\delta_{\mathrm{GL}}(y^{Mn}) = \arg\max_i \log \hat{\hat{p}}_i(y^{Mn})$$

$$= \arg\min_i \sum_{j\neq i} D\left(\gamma_j \,\Big\|\, \frac{1}{M-1}\sum_{k\neq i}\gamma_k\right)$$

Proposed Universal Test

(μ, π) not known: $\hat{\mu}_i = \gamma_i \qquad \hat{\pi}_i = \frac{1}{M-1} \sum_{j \neq i} \gamma_j$

$$H_i: \quad \hat{\hat{p}}_i \left(y^{Mn} \right) = \prod_{k=1}^{n} \left[\hat{\mu}_i \left(y_k^{(i)} \right) \prod_{j \neq i} \hat{\pi}_i \left(y_k^{(j)} \right) \right]$$

$$\begin{array}{c|ccccc} & 1 & 2 & & & M \\ \hline H_1 & \mu & \pi & \pi & \pi & \pi \\ H_2 & \pi & \mu & \pi & \pi & \pi \\ & \pi & \pi & \mu & \pi & \pi \\ & \pi & \pi & \pi & \mu & \pi \\ H_M & \pi & \pi & \pi & \pi & \mu \end{array}$$

Generalized Likelihood (GL) Rule:

$$\delta_{\mathrm{GL}}(y^{Mn}) = \arg\max_i \log \hat{\hat{p}}_i (y^{Mn})$$

$$= \arg\min_i \sum_{j \neq i} D\left(\gamma_j \,\Big\|\, \frac{1}{M-1} \sum_{k \neq i} \gamma_k \right)$$

$\longleftarrow$ key statistic

Performance of Universal Test

$$\alpha\Big(\delta, \big(\mu, \pi\big)\Big) = \min_{q_1, \ldots, q_M} D(q_1 \,||\, \mu) + D(q_2 \,||\, \pi) + \ldots + D(q_M \,||\, \pi)$$

$$\sum_{j \neq 1} D(q_j \,||\, \frac{1}{M-1} \sum_{k \neq 1} q_k) \;\; \geq \;\; \sum_{j \neq 2} D(q_j \,||\, \frac{1}{M-1} \sum_{k \neq 2} q_k)$$

Performance of Universal Test

Universally exponential consistency!

$$\alpha\left(\delta, \left(\mu, \pi\right)\right) = \min_{q_1, \ldots, q_M} D(q_1 \,||\, \mu) + D(q_2 \,||\, \pi) + \ldots + D(q_M \,||\, \pi)$$

$$> 0, \ \forall \left(\mu, \pi\right)$$

$$\sum_{j \neq 1} D(q_j \,||\, \frac{1}{M-1} \sum_{k \neq 1} q_k) \ \geq \ \sum_{j \neq 2} D(q_j \,||\, \frac{1}{M-1} \sum_{k \neq 2} q_k)$$

Asymptotic Efficiency

- Motivation: When only π is known, optimal error exponent is $2B\left(\mu, \pi\right)$

- Estimate of π satisfies

$$\lim_{n\to\infty} \frac{1}{M}\sum_{i=1}^{M}\gamma_i = \frac{1}{M}\mu + \frac{M-1}{M}\pi,$$

$$\lim_{M\to\infty} \frac{1}{M}\mu + \frac{M-1}{M}\pi = \pi$$

Asymptotic Efficiency

Our universal outlier detector achieves error exponent lower bounded by

$$\min_{q:\, D(q\|\pi) \leq \frac{1}{M-1}(\,2B(\mu,\,\pi)+C_{\pi}\,)} 2B(\mu,\,q)$$

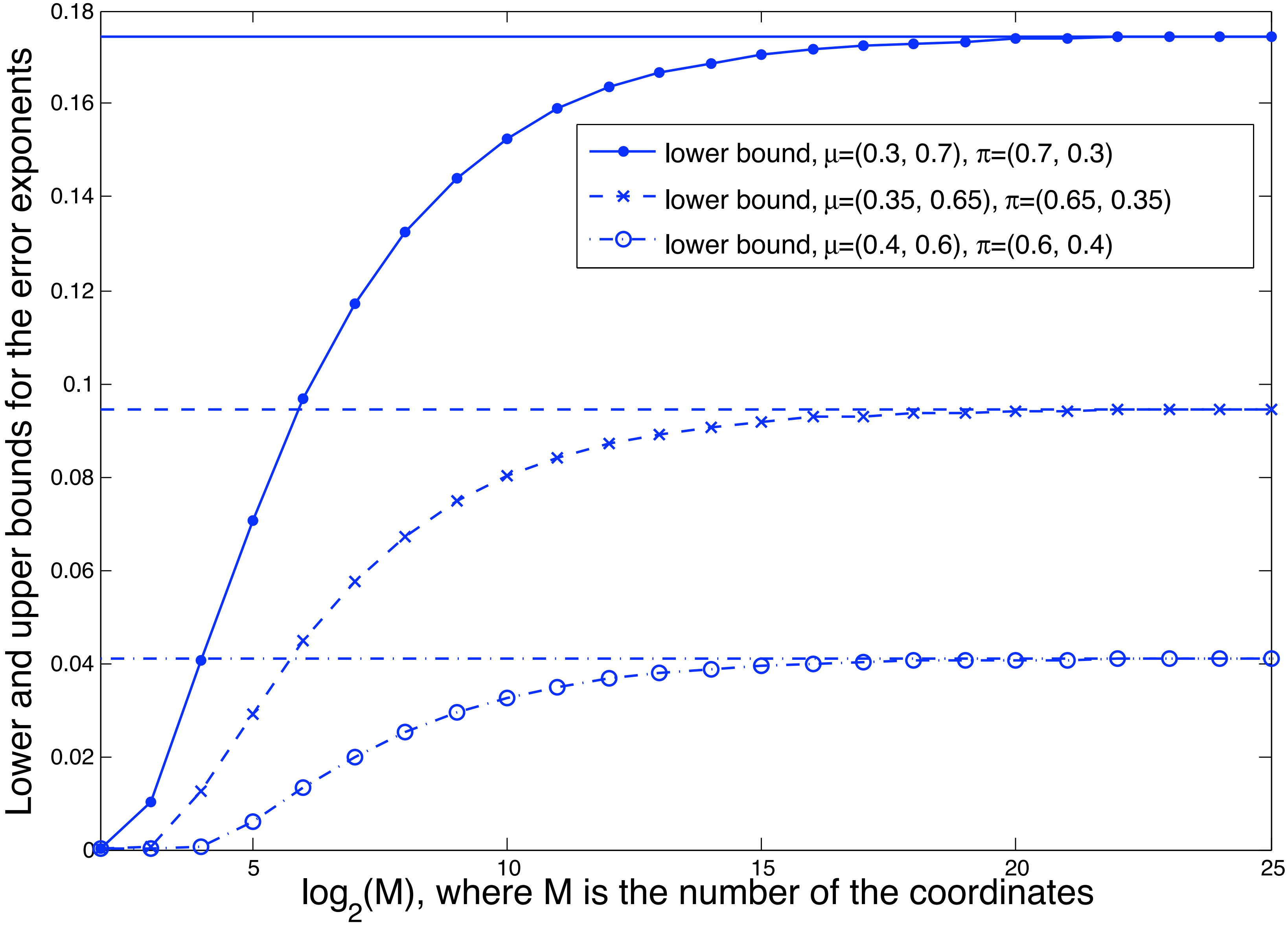
Asymptotic Efficiency

Our universal outlier detector achieves error exponent lower bounded by

$$\min_{q:\ D(q\|\pi) \leq \frac{1}{M-1}\left(2B(\mu,\ \pi)+C_\pi\right)} 2B(\mu,\ q)$$

This lower bound is non-decreasing in $M \geq 3$, and converges to $2B(\mu,\ \pi)$ as $M \to \infty$

Numerical Results



Extension to Multiple Outliers: Known (μ,π)

For $S \subseteq \{1, \ldots, M\}, \ |S| = T, \ T$ fixed and known

$$H_S : \quad p_S\left(\mathbf{y}^{Mn}\right) = \prod_{k=1}^{n}\left[\prod_{i \in S} \mu\left(y_k^{(i)}\right) \prod_{j \notin S} \pi\left(y_k^{(j)}\right)\right]$$

ML Rule : $\quad \delta_{\mathrm{ML}}(y^{Mn}) = \arg\max_S \log p_S(y^{Mn})$

Exponential Consistency : $\quad \alpha(\delta_{\mathrm{ML}},(\mu,\pi)) = 2B(\mu,\pi)$

Proposed Universal Test: One Outlier

Generalized Likelihood (GL) Rule :

$$\begin{aligned}\delta_{\mathrm{GL}}(y^{Mn}) &= \arg\max_{i} \log \hat{\hat{p}}_i(y^{Mn}) \\ &= \arg\min_{i} \sum_{j \neq i} D\left(\gamma_j \,\Big\|\, \frac{1}{M-1}\sum_{k \neq i} \gamma_k\right) \longleftarrow \text{key statistic}\end{aligned}$$

$$\begin{array}{c|ccccc|} & 1 & 2 & & & M \\ \hline H_1 & \mu & \pi & \pi & \pi & \pi \\ H_2 & \pi & \mu & \pi & \pi & \pi \\ & \pi & \pi & \mu & \pi & \pi \\ & \pi & \pi & \pi & \mu & \pi \\ H_M & \pi & \pi & \pi & \pi & \mu \\ \hline \end{array}$$

Proposed Universal Test: One Outlier

Generalized Likelihood (GL) Rule :

$$
\begin{aligned}
\delta_{\mathrm{GL}}(y^{Mn}) &= \arg\max_i \log \hat{\hat{p}}_i(y^{Mn}) \\
&= \arg\min_i \sum_{j \neq i} D\left(\gamma_j \,\middle\|\, \frac{1}{M-1}\sum_{k \neq i}\gamma_k\right)
\end{aligned}
$$

key statistic

Summing over all typical sequence indices under hypothesis i

$\hat{\pi}_i$

$$
\begin{array}{c|ccccc|}
 & 1 & 2 & & & M \\
\hline
H_1 & \mu & \pi & \pi & \pi & \pi \\
H_2 & \pi & \mu & \pi & \pi & \pi \\
 & \pi & \pi & \mu & \pi & \pi \\
 & \pi & \pi & \pi & \mu & \pi \\
H_M & \pi & \pi & \pi & \pi & \mu \\
\hline
\end{array}
$$

Proposed Universal Test: Multiple Outliers

Generalized Likelihood (GL) Rule :

$$\delta_{\mathrm{GL}}(y^{Mn}) = \underset{S}{\arg\min} \sum_{j \notin S} D\left(\gamma_j \,\Big\|\, \underbrace{\frac{1}{M-T} \sum_{k \notin S} \gamma_k}_{\hat{\pi}_S}\right)$$

Summing over all typical sequence indices under hypothesis S

$\hat{\pi}_S$

Asymptotic Efficiency

Our universal outlier detector achieves error exponent lower bounded by

$$\min_{q:\ D(q\|\pi) \le \frac{1}{M-T}\left(2B(\mu,\pi)+C_\pi\right)} 2B(\mu, q)$$

Asymptotic Efficiency

Our universal outlier detector achieves error exponent lower bounded by

$$\min_{q:\, D(q\|\pi) \le \frac{1}{M-T}\left(2B(\mu,\pi)+C_\pi\right)} 2B(\mu, q)$$

This lower bound is non-decreasing in M, and converges to $2B(\mu, \pi)$ as $M \to \infty$

Conclusion

- Generalized likelihood (GL) test is universally exponentially consistent for outlier hypothesis testing wherein number of outliers is fixed and known *a priori*
- GL test is asymptotically efficient in error exponent for large *M* even with no training data
- If number of outliers is not known *a priori*, there is no universally exponentially consistent test

Fraud Detection: Group Monitoring

- Male graduate students

Student 1	Student 2	Student 3	…	Student *M*
Grocery	Dining	Grocery	…	Gas
Dining	Grocery	Gas	…	Books
…	…	…	…	…
Books	Movie	Books	…	Grocery
Movie	Books	Grocery	…	Dining
Grocery	Gas	spa	…	Movie
Gas	Books	Cosmetics	…	Grocery