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

<table>
<thead>
<tr>
<th>Transactions</th>
<th>Grocery</th>
<th>Gas</th>
<th>---</th>
<th>Books</th>
<th>Grocery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amount</td>
<td>30$</td>
<td>35$</td>
<td>---</td>
<td>75$</td>
<td>35$</td>
</tr>
</tbody>
</table>

Generic behavior
### Fraud Detection

- **Example:** spending records for a male graduate student

<table>
<thead>
<tr>
<th>Transactions</th>
<th>Grocery</th>
<th>Gas</th>
<th>---</th>
<th>Books</th>
<th>Grocery</th>
<th>Spa</th>
<th>Cosmetics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amount</td>
<td>30$</td>
<td>35$</td>
<td>---</td>
<td>75$</td>
<td>35$</td>
<td>250$</td>
<td>500$</td>
</tr>
</tbody>
</table>

**Generic behavior**

**Fraudulent behavior**
Fraud Detection: Group Monitoring

- Male graduate students

<table>
<thead>
<tr>
<th>Student 1</th>
<th>Student 2</th>
<th>Student 3</th>
<th>...</th>
<th>Student M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grocery</td>
<td>Dining</td>
<td>Grocery</td>
<td>...</td>
<td>Gas</td>
</tr>
<tr>
<td>Dining</td>
<td>Grocery</td>
<td>Gas</td>
<td>...</td>
<td>Books</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Books</td>
<td>Movie</td>
<td>Books</td>
<td>...</td>
<td>Grocery</td>
</tr>
<tr>
<td>Movie</td>
<td>Books</td>
<td>Grocery</td>
<td>...</td>
<td>Dining</td>
</tr>
<tr>
<td>Grocery</td>
<td>Gas</td>
<td>spa</td>
<td>...</td>
<td>Movie</td>
</tr>
<tr>
<td>Gas</td>
<td>Books</td>
<td>Cosmetics</td>
<td>...</td>
<td>Grocery</td>
</tr>
</tbody>
</table>
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 $\pi$
- Outlier distribution $\mu$

```
\begin{array}{cccc}
  & 1 & 2 & M \\
H_1 & \mu & \pi & \pi & \pi & \pi \\
H_2 & \pi & \mu & \pi & \pi & \pi \\
H_M & \pi & \pi & \pi & \mu & \pi \\
\end{array}
```
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

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

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

Nothing is known about \((\mu, \pi)\) except that they are distinct

Universal Test: \[ \delta : \cal{Y}^{\text{Mn}} \rightarrow \{1, \ldots, M\} \]
Universal Outlier Hypothesis Testing

\[ H_i : p_i(y^{Mn}) = \prod_{k=1}^{n} \mu(y^{(i)}_k) \prod_{j \neq i} \pi(y^{(j)}_k) \]

Nothing is known about \((\mu, \pi)\) except that they are distinct

Universal Test: \( \delta : \mathcal{Y}^{Mn} \to \{1, \ldots, M\} \)

Independent of \((\mu, \pi)\)
Performance Metrics

• Maximal error probability:

\[ e(\delta, (\mu, \pi)) = \max_i \mathbb{P}_i \{ \delta(y^{Mn}) \neq i \} \]

• Exponent for maximal error probability:

\[ \alpha(\delta, (\mu, \pi)) = \lim_{n \to \infty} -\frac{1}{n} \log e(\delta, (\mu, \pi)) \]
Performance Metrics

• Maximal error probability:

\[ e(\delta, (\mu, \pi)) = \max_i \mathbb{P}_i \{ \delta(y^{Mn}) \neq i \} \]

• Exponent for maximal error probability:

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

Consistency: \( e \to 0 \) as \( n \to \infty \)

Exponential Consistency: \( \alpha > 0 \)
Background: Binary Hypothesis Testing

\[ H_1 : p_1(y) = \prod_{k=1}^{n} \pi(y_k) \quad H_2 : p_2(y) = \prod_{k=1}^{n} \mu(y_k) \]

If \((\mu, \pi)\) known, \(\delta_{\text{ML}}(y) = \arg\max_i \log p_i(y)\)

has \(\alpha(\delta, (\mu, \pi)) = C(\mu, \pi) > 0\)
**Background: Binary Hypothesis Testing**

\[
H_1 : p_1(y) = \prod_{k=1}^{n} \pi(y_k) \quad H_2 : p_2(y) = \prod_{k=1}^{n} \mu(y_k)
\]

If \((\mu, \pi)\) known, \(\delta_{\text{ML}}(y) = \arg\max_i \log p_i(y)\) has \(\alpha(\delta, (\mu, \pi)) = C(\mu, \pi) > 0\) \(\leftarrow\) exponential consistency
Background: Binary Hypothesis Testing

\[ H_1: p_1(y) = \prod_{k=1}^{n} \pi(y_k) \quad H_2: p_2(y) = \prod_{k=1}^{n} \mu(y_k) \]

If \((\mu, \pi)\) known, \(\delta_{\text{ML}}(y) = \arg\max_i \log p_i(y)\)

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

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

ML Rule : \[\delta_{ML}(y^{Mn}) = \arg \max_i \log p_i(y^{Mn})\]

Exponential Consistency : \[\alpha(\delta_{ML},(\mu, \pi)) = 2B(\mu, \pi)\]

Bhattacharyya Distance : \[B(\mu, \pi) = -\log \left( \sum_y \mu(y)^{1/2} \pi(y)^{1/2} \right)\]
Binary Hypothesis Testing: Unknown $\mu$

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

If $\mu$ unknown

for any given $\delta$ there exists $\mu$ s.t. $\alpha = 0$

No exponential consistency!
μ unknown: \( \hat{\mu}_i = \gamma_i \leftarrow \) empirical distribution

\[ H_i : \hat{\rho}_i(y^{Mn}) = \prod_{k=1}^{n} \begin{bmatrix} \hat{\mu}_i(y_k^{(i)}) \prod_{j \neq i} \pi(y_k^{(j)}) \end{bmatrix} \]

Generalized Likelihood (GL) Rule:
\[ \delta_{GL}(y^{Mn}) = \arg\max_i \log \hat{\rho}_i(y^{Mn}) \]

Exponential Consistency: \( \alpha(\delta_{GL},(\mu,\pi)) = 2B(\mu,\pi) \)
Outlier Hypothesis Testing: Unknown $\mu$

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

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

Generalized Likelihood (GL) Rule:

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

Exponential Consistency:

$\alpha(\delta_{GL}, (\mu, \pi)) = 2B(\mu, \pi)$

Same as known $\mu, \pi$
Sanov’s Theorem: For i.i.d. rvs $Y^n \sim \rho$, 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 $Y^n \sim \rho$, exponent of probability that random empirical distribution falls in closed set $E$ is

$$
\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left\{ \text{Empirical}(Y^n) \in E \right\} = \min_{q \in E} D(q \| \rho)
$$
Proposed Universal Test

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

\[H_i : \hat{\rho}_i(y^M) = \prod_{k=1}^{n} \hat{\mu}_i(y_k^{(i)}) \prod_{j \neq i} \hat{\pi}_i(y_k^{(j)})\]

Generalized Likelihood (GL) Rule:
\[\delta_{GL}(y^M) = \arg \max_i \log \hat{\rho}_i(y^M)\]
Proposed Universal Test

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

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

Generalized Likelihood (GL) Rule:

\[ \delta_{GL}(\mathbf{y}^M) = \arg\max_{i} \log \hat{p}_i(\mathbf{y}^M) \]

\[ = \arg\min_{i} \sum_{j \neq i} D \left( \gamma_j \parallel \frac{1}{M - 1} \sum_{k \neq i} \gamma_k \right) \]
Proposed Universal Test

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

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

Generalized Likelihood (GL) Rule:

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

\[= \arg\min_i \sum_{j \neq i} D\left(\gamma_j \parallel \frac{1}{M-1} \sum_{k \neq i} \gamma_k\right)\]
Performance of Universal Test

$$\alpha(\delta, (\mu, \pi)) = \min_{q_1, \ldots, q_M} \sum_{j \neq 1} D(q_j \parallel \mu) + D(q_2 \parallel \pi) + \ldots + D(q_M \parallel \pi)$$

$$\sum_{j \neq 1} D(q_j \parallel \frac{1}{M-1} \sum_{k \neq 1} q_k) \geq \sum_{j \neq 2} D(q_j \parallel \frac{1}{M-1} \sum_{k \neq 2} q_k)$$
Universally exponential consistency!

\[ \alpha(\delta, (\mu, \pi)) = \min_{q_1, \ldots, q_M} D(q_1 \| \mu) + D(q_2 \| \pi) + \ldots + D(q_M \| \pi) \]

\[ > 0, \quad \forall (\mu, \pi) \]

\[ \sum_{j \neq 1} D(q_j \| \frac{1}{M-1} \sum_{k \neq 1} q_k) \geq \sum_{j \neq 2} D(q_j \| \frac{1}{M-1} \sum_{k \neq 2} q_k) \]
Asymptotic Efficiency

• Motivation: When only $\pi$ is known, optimal error exponent is $2B(\mu, \pi)$

• Estimate of $\pi$ satisfies

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

\[
\lim_{M \to \infty} \frac{1}{M} \mu + \frac{M - 1}{M} \pi = \pi
\]
Our universal outlier detector achieves error exponent lower bounded by

\[
\min_{q: D(q\|\pi) \leq \frac{1}{M-1}(2B(\mu, \pi) + C_\pi)} 2B(\mu, q)
\]
Our universal outlier detector achieves error exponent lower bounded by

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

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

Lower and upper bounds for the error exponents

\[ \log_2(M), \text{ where } M \text{ is the number of the coordinates} \]

- Lower bound, \( \mu=(0.3, 0.7), \pi=(0.7, 0.3) \)
- Lower bound, \( \mu=(0.35, 0.65), \pi=(0.65, 0.35) \)
- Lower bound, \( \mu=(0.4, 0.6), \pi=(0.6, 0.4) \)
Extension to Multiple Outliers: Known \((\mu, \pi)\)

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

\[
H_S : \quad p_S(y^{Mn}) = \prod_{k=1}^{n} \left[ \prod_{i \in S} \mu(y^{(i)}_k) \prod_{j \not\in S} \pi(y^{(j)}_k) \right]
\]

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

Exponential Consistency: \(\alpha(\delta_{ML},(\mu, \pi)) = 2B(\mu, \pi)\)
### Proposed Universal Test: One Outlier

#### Generalized Likelihood (GL) Rule:

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

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

---

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th></th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_1$</td>
<td>$\mu$</td>
<td>$\pi$</td>
<td>$\pi$</td>
<td>$\pi$</td>
</tr>
<tr>
<td>$H_2$</td>
<td>$\pi$</td>
<td>$\mu$</td>
<td>$\pi$</td>
<td>$\pi$</td>
</tr>
<tr>
<td>$H_3$</td>
<td>$\pi$</td>
<td>$\pi$</td>
<td>$\mu$</td>
<td>$\pi$</td>
</tr>
<tr>
<td>$H_M$</td>
<td>$\pi$</td>
<td>$\pi$</td>
<td>$\pi$</td>
<td>$\mu$</td>
</tr>
</tbody>
</table>
Proposed Universal Test: One Outlier

Generalized Likelihood (GL) Rule:

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

\[
= \arg\min_i \sum_{j \neq i} D \left( \gamma_j \parallel \frac{1}{M-1} \sum_{k \neq i} \gamma_k \right)
\]

Summing over all typical sequence indices under hypothesis \(i\)

Key statistic
Proposed Universal Test: Multiple Outliers

Generalized Likelihood (GL) Rule:

$$\delta_{GL}(y^{Mn}) = \arg\min_S \sum_{j \notin S} D \left( \gamma_j \parallel \frac{1}{M - T} \sum_{k \notin S} \gamma_k \right)$$

Summing over all typical sequence indices under hypothesis $S$
Asymptotic Efficiency

Our universal outlier detector achieves error exponent lower bounded by

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

\[
2B(\mu, q)
\]
Our universal outlier detector achieves error exponent lower bounded by

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

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 \textit{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 \textit{a priori}, there is no universally exponentially consistent test.
Fraud Detection: Group Monitoring

- Male graduate students

<table>
<thead>
<tr>
<th>Student 1</th>
<th>Student 2</th>
<th>Student 3</th>
<th>…</th>
<th>Student M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grocery</td>
<td>Dining</td>
<td>Grocery</td>
<td>…</td>
<td>Gas</td>
</tr>
<tr>
<td>Dining</td>
<td>Grocery</td>
<td>Gas</td>
<td>…</td>
<td>Books</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>Books</td>
<td>Movie</td>
<td>Books</td>
<td>…</td>
<td>Grocery</td>
</tr>
<tr>
<td>Movie</td>
<td>Books</td>
<td>Grocery</td>
<td>…</td>
<td>Dining</td>
</tr>
<tr>
<td>Grocery</td>
<td>Gas</td>
<td>spa</td>
<td>…</td>
<td>Movie</td>
</tr>
<tr>
<td>Gas</td>
<td>Books</td>
<td>Cosmetics</td>
<td>…</td>
<td>Grocery</td>
</tr>
</tbody>
</table>