Universal Outlier Hypothesis Testing

Venu Veeravalli
ECE Dept & CSL & ITI
University of Illinois at Urbana-Champaign
http://www.ifp.illinois.edu/~vvv

(with Sirin Nitinawarat and Yun Li)

NSF Workshop on Signal Processing for Big Data
March 22, 2013
Statistical Outlier Detection

• Single sequence of observations

• Normal 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 sensor data, etc.
Fraud Detection

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

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

Normal behavior
**Example:** spending records for a male graduate student

<table>
<thead>
<tr>
<th>Transactions</th>
<th>Grocery</th>
<th>Gas</th>
<th>Books</th>
<th>---</th>
<th>Spa</th>
<th>Women apparels</th>
<th>Cosmetics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amount</td>
<td>30$</td>
<td>35$</td>
<td>75$</td>
<td></td>
<td>250$</td>
<td>1500$</td>
<td>500$</td>
</tr>
</tbody>
</table>

Normal behavior

Abnormal behavior

Veeravalli, NSF Big Data 3/22/13
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>Books</td>
<td>Movie</td>
<td>Books</td>
<td>...</td>
<td>Grocery</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Movie</td>
<td>Books</td>
<td>Spa</td>
<td>...</td>
<td>Dining</td>
</tr>
<tr>
<td>Grocery</td>
<td>Gas</td>
<td>Women apparel</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) distributions
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) distributions
• Special case:
  o Exactly one sequence is generated from outlier distribution
  o Goal: to detect outlier sequence efficiently
  o Universal setting: neither typical nor outlier distributions known; no training data provided
### Universal Outlier Hypothesis Testing

- **Typical distribution** $\pi$
- **Outlier distribution** $\mu$

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_1$</td>
<td>$\mu$</td>
<td>$\pi$</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>
<td>$\pi$</td>
</tr>
<tr>
<td>$H_M$</td>
<td>$\pi$</td>
<td>$\pi$</td>
<td>$\pi$</td>
<td>$\mu$</td>
<td>$\pi$</td>
</tr>
</tbody>
</table>

Veeravalli, NSF Big Data 3/22/13
Mathematical Model

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

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

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

Universal Detector: \[ \delta : \mathcal{Y}^{Mn} \rightarrow \{1, \ldots, M\} \]

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

• Maximal error probability:

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

• 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 \(\alpha(\delta, (\mu, \pi)) = C(\mu, \pi) > 0 \quad \leftarrow\) 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 \((\mu, \pi)\)

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

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 $\mu$

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

If $\mu$ unknown
for any given $\delta$ there exists $\mu$ s.t. $\alpha = 0$
No exponential consistency!
Outlier Hypothesis Testing: unknown $\mu$

$\mu$ unknown: $\hat{\mu}_i = \gamma_i$ ← empirical Distribution

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

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 $\mu, \pi$

Veeravalli, NSF Big Data 3/22/13
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

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

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

Empirical distributions

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

Generalized Likelihood (GL) Rule:

\[ \delta_{GL}(y^{Mn}) = \arg \max_i \log \hat{\rho}_i(y^{Mn}) \]
Performance of Universal Test

\[
\alpha\left(\delta, \left(\mu, \pi\right)\right) = \min_{p_1, \ldots, p_M} D(p_1 \parallel \mu) + D(p_2 \parallel \pi) + \ldots + D(p_M \parallel \pi)
\]

\[
\sum_{j \neq 1} D(p_j \parallel \frac{1}{M-1} \sum_{k \neq 1} p_k) \geq \sum_{j \neq 2} D(p_j \parallel \frac{1}{M-1} \sum_{k \neq 2} p_k)
\]
Universally exponential consistency!

\[ \alpha(\delta, (\mu, \pi)) = \min_{p_1, \ldots, p_M} D(p_1 \| \mu) + D(p_2 \| \pi) + \ldots + D(p_M \| \pi) \]

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

\[ \sum_{j \neq 1} D(p_j \| \frac{1}{M-1} \sum_{k \neq 1} p_k) \geq \sum_{j \neq 2} D(p_j \| \frac{1}{M-1} \sum_{k \neq 2} p_k) \]
Asymptotic Optimality

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

• Estimate of $\pi$ satisfies

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

\[
\lim_{M \to \infty} \frac{1}{M} \mu + \frac{M - 1}{M} \pi = \pi
\]
Asymptotic Optimality

Our universal outlier detector achieves error exponent lower bounded by

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

Veeravalli, NSF Big Data 3/22/13
Our universal outlier detector achieves error exponent lower bounded by

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

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

Lower bound of the error exponent

\[ 2 B (\mu, \pi), \mu=(0.36, 0.64) \]
\[ \pi=(0.64, 0.36) \]
Discussion

• Generalized likelihood (GL) test is universally exponentially consistent for single outlier hypothesis testing
• GL test is asymptotically optimum in error exponent for large $M$
• What if we have more than one outlier?
Discussion

• Generalized likelihood (GL) test is universally exponentially consistent for single outlier hypothesis testing

• GL test is asymptotically optimum in error exponent for large $M$

• What if we have more than one outlier?
  
  o If it is known that we have exactly $K << M$ outliers, then GL test is still universally exponentially consistent
• Generalized likelihood (GL) test is universally exponentially consistent for single outlier hypothesis testing
• GL test is asymptotically optimum in error exponent for large $M$
• What if we have more than one outlier?
  o If it is known that we have exactly $K << M$ outliers, then GL test is still universally exponentially consistent
  o If number of outliers is not known a priori, then universally exponentially does not exist!