Deviation Detection with Continuous Observations

Pengfei Yang and Biao Chen

Syracuse University

pyang04{bichen}@syr.edu

December 11, 2015

Overview

1. Deviation Detection Problem Formulation
2. Why Use d_L Rather Than d_{KL}.
3. Asymptotic Deviation Detection
   - An Asymptotically $\delta-$Optimal Detector.

Overview

1. Deviation Detection Problem Formulation
2. Why Use d_L Rather Than d_{KL}.
3. Asymptotic Deviation Detection
   - An Asymptotically $\delta-$Optimal Detector.

General Model

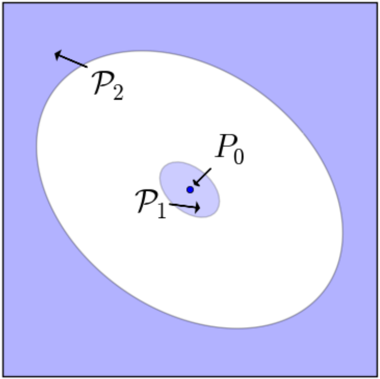

Fig. 1: the probability space $\mathcal{P}$.

P_0 : the nominal distribution.
$\mathcal{P}$: the probability space.

$$\mathcal{P}_1 := \{P \in \mathcal{P} : d(P, P_0) \leq \lambda_1\},$$
$$\mathcal{P}_2 := \{P \in \mathcal{P} : d(P, P_0) \geq \lambda_2\}.$$

General Model

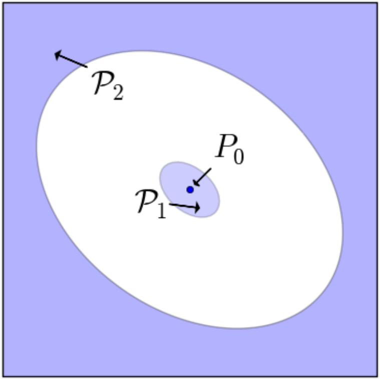

Fig. 1: the probability space $\mathcal{P}$.

P_0 : the nominal distribution.
$\mathcal{P}$: the probability space.

$$\mathcal{P}_1 := \{P \in \mathcal{P} : d(P, P_0) \leq \lambda_1\},$$
$$\mathcal{P}_2 := \{P \in \mathcal{P} : d(P, P_0) \geq \lambda_2\}.$$

$$\mathcal{H}_1 : P \in \mathcal{P}_1,$$
$$\mathcal{H}_2 : P \in \mathcal{P}_2.$$

Under Neyman-Pearson (N-P) criterion, which d is appropriate?

Discrete Case

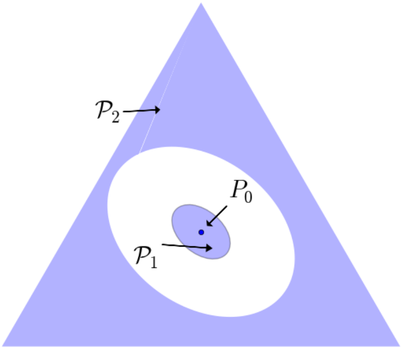

Fig. 2: the probability simplex $\mathcal{P}$.

P_0 : a discrete distribution.

$$\mathcal{P}_1 := \{P \in \mathcal{P} : d(P, P_0) \leq \lambda_1\},$$
$$\mathcal{P}_2 := \{P \in \mathcal{P} : d(P, P_0) \geq \lambda_2\}.$$

Discrete Case

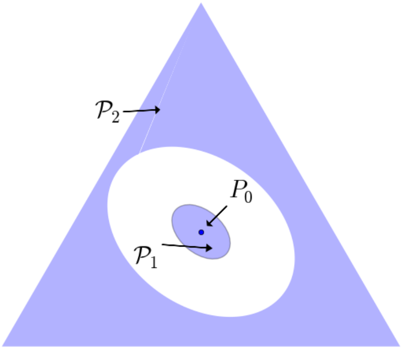

Fig. 2: the probability simplex $\mathcal{P}$.

P_0 : a discrete distribution.

$$\mathcal{P}_1 := \{P \in \mathcal{P} : d(P, P_0) \leq \lambda_1\},$$
$$\mathcal{P}_2 := \{P \in \mathcal{P} : d(P, P_0) \geq \lambda_2\}.$$

Many choices for d, e.g.,
the total variation,
KL divergence ...

Continuous Case

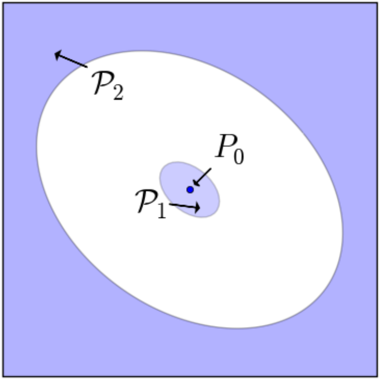

Fig. 3: the probability space $\mathcal{P}$.

P_0 : the nominal distribution.
$\mathcal{P}$: the probability space.

$$\mathcal{P}_1 := \{P \in \mathcal{P} : d(P, P_0) \leq \lambda_1\},$$
$$\mathcal{P}_2 := \{P \in \mathcal{P} : d(P, P_0) \geq \lambda_2\}.$$

$$\mathcal{H}_1 : P \in \mathcal{P}_1,$$
$$\mathcal{H}_2 : P \in \mathcal{P}_2.$$

Unlike the discrete case, not every d can describe such a problem.

Composite Hypothesis Testing

The minimax N-P criterion [Huber, 1965]

$$\min_{\phi^n} \sup_{P_2 \in \mathcal{P}_2} P_2^n(\phi^n = 1) \quad \text{s.t.} \quad \sup_{P_1 \in \mathcal{P}_1} P_1^n(\phi^n = 2) \leq \alpha,$$

where $\phi^n = \phi^n(X^n)$ is the output of the detector.

Composite Hypothesis Testing

The minimax N-P criterion [Huber, 1965]

$$\min_{\phi^n} \sup_{P_2 \in \mathcal{P}_2} P_2^n(\phi^n = 1) \quad \text{s.t.} \quad \sup_{P_1 \in \mathcal{P}_1} P_1^n(\phi^n = 2) \leq \alpha,$$

where $\phi^n = \phi^n(X^n)$ is the output of the detector.

Natrually, we need $\mathcal{P}_1 \cap \mathcal{P}_2 = \emptyset$.

Composite Hypothesis Testing

The minimax N-P criterion [Huber, 1965]

$$\min_{\phi^n} \sup_{P_2 \in \mathcal{P}_2} P_2^n(\phi^n = 1) \quad \text{s.t.} \quad \sup_{P_1 \in \mathcal{P}_1} P_1^n(\phi^n = 2) \leq \alpha,$$

where $\phi^n = \phi^n(X^n)$ is the output of the detector.

Natrually, we need $\mathcal{P}_1 \cap \mathcal{P}_2 = \emptyset$.

Actually, we need $cl\mathcal{P}_1 \cap cl\mathcal{P}_2 = \emptyset$, where cl is with respect to $(\mathcal{P}, d_L)$.

Composite Hypothesis Testing

The minimax N-P criterion [Huber, 1965]

$$\min_{\phi^n} \sup_{P_2 \in \mathcal{P}_2} P_2^n(\phi^n = 1) \quad \text{s.t.} \quad \sup_{P_1 \in \mathcal{P}_1} P_1^n(\phi^n = 2) \leq \alpha,$$

where $\phi^n = \phi^n(X^n)$ is the output of the detector.

Natrually, we need $\mathcal{P}_1 \cap \mathcal{P}_2 = \emptyset$.

Actually, we need $cl\mathcal{P}_1 \cap cl\mathcal{P}_2 = \emptyset$, where cl is with respect to $(\mathcal{P}, d_L)$.

$d_L(F, G) := \inf\{\epsilon : F(x - \epsilon) - \epsilon \leq G(x) \leq F(x + \epsilon) + \epsilon, \forall x\}$.

Overview

Closure of The KL Surface

Proposition 1.

Let P_0 be the normal distribution. For any given $\lambda > 0$, let

$$\mathcal{P}_\lambda := \{P \in \mathcal{P} : d_{KL}(P, P_0) = \lambda\},$$

then

$$cl\mathcal{P}_\lambda = \{P \in \mathcal{P} : d_{KL}(P, P_0) \leq \lambda\}.$$

Closure of The KL Surface

Proposition 1.

Let P_0 be the normal distribution. For any given $\lambda > 0$, let

$$\mathcal{P}_\lambda := \{P \in \mathcal{P} : d_{KL}(P, P_0) = \lambda\},$$

then

$$cl\mathcal{P}_\lambda = \{P \in \mathcal{P} : d_{KL}(P, P_0) \leq \lambda\}.$$

If d_{KL} is used in defining $\mathcal{P}_1$ and $\mathcal{P}_2$, then,

Closure of The KL Surface

Proposition 1.

Let P_0 be the normal distribution. For any given $\lambda > 0$, let

$$\mathcal{P}_\lambda := \{P \in \mathcal{P} : d_{KL}(P, P_0) = \lambda\},$$

then

$$cl\mathcal{P}_\lambda = \{P \in \mathcal{P} : d_{KL}(P, P_0) \leq \lambda\}.$$

If d_{KL} is used in defining $\mathcal{P}_1$ and $\mathcal{P}_2$, then,

- $cl\mathcal{P}_1 \cap cl\mathcal{P}_2 \neq \emptyset$.

d_{KL} is not appropriate in defining the deviation detection problem.

The Legitimacy of Choosing d_L

By defining $\mathcal{P}_1$ and $\mathcal{P}_2$ using d_L,

The Legitimacy of Choosing d_L

By defining $\mathcal{P}_1$ and $\mathcal{P}_2$ using d_L,

- $cl\mathcal{P}_1 \cap cl\mathcal{P}_2 = \emptyset$,
- The proximity set $\mathcal{P}_1$ inlcudes *all distributions which are close enough to P_0 in other measures, for example d_{KL}.*

The Legitimacy of Choosing d_L

By defining $\mathcal{P}_1$ and $\mathcal{P}_2$ using d_L,

- $cl\mathcal{P}_1 \cap cl\mathcal{P}_2 = \emptyset$,
- The proximity set $\mathcal{P}_1$ inlcudes *all distributions which are close enough to P_0 in other measures, for example d_{KL}.*
  - *For any $\lambda_1' > 0$, there exists a $\lambda_1 > 0$ s.t.*

$$B_{KL}(P_0, \lambda_1) \subseteq B_L(P_0, \lambda_1'),$$

  - *however, the reverse statement is not true, i.e., for any λ_1, λ_1',*

$$B_L(P_0, \lambda_1') \nsubseteq B_{KL}(P_0, \lambda_1).$$

Overview

Asymptotic deviation detection

For the fixed sample size deviation detection, some widely used methods fail, e.g.,

- generalized likelihood ratio test (GLRT),
- least favorable distributions (LFDs).

Asymptotic deviation detection

For the fixed sample size deviation detection, some widely used methods fail, e.g.,

- generalized likelihood ratio test (GLRT),
- least favorable distributions (LFDs).

But we can still find an asymptotically $\delta-$optimal detector.

Asymptotic deviation detection

$\hat{\mu}_n \sim$ the empirical distribution.

The minimax asymptotic N-P criterion. [Zeitouni et al., 1991]

$$\sup_{\Omega} \inf_{P_2 \in \mathcal{P}_2} I^{P_2}(\Omega) \quad \text{s.t.} \quad \inf_{P_1 \in \mathcal{P}_1} J^{P_1}(\Omega) \geq \eta,$$

where

$$I^{P_2}(\Omega) = \varliminf_{n \to \infty} -\frac{1}{n} \log P_2^n(\hat{\mu}_n \in \Omega_1(n)),$$

$$J^{P_1}(\Omega) = \varliminf_{n \to \infty} -\frac{1}{n} \log P_1^n(\hat{\mu}_n \in \Omega_2(n)).$$

$\Omega = \{(\Omega_1(n), \Omega_2(n)), n \geq 1\}$ in which $\Omega_1(n)$ and $\Omega_2(n)$ are partitions of all n-sample empirical distributions.

Asymptotic deviation detection

1. When $\mathcal{P}_1$ and $\mathcal{P}_2$ each contains only single continuous distribution, [Zeitouni et al., IEEE Trans IT 1991] provides an $\delta-$optimal detector.
2. When $\mathcal{P}_1$ represents moment restrictions, [Kitamura, 2001] demonstrated that $\delta-$optimality holds for the empirical likelihood ratio test.
3. In our problem, $\mathcal{P}_1 := \{P \in \mathcal{P} : d_L(P, P_0) \leq \lambda_1\}, \mathcal{P}_2 := \{P \in \mathcal{P} : d_L(P, P_0) \geq \lambda_2\}$.

Asymptotic deviation detection

1. When $\mathcal{P}_1$ and $\mathcal{P}_2$ each contains only single continuous distribution, [Zeitouni et al., IEEE Trans IT 1991] provides an $\delta-$optimal detector.
2. When $\mathcal{P}_1$ represents moment restrictions, [Kitamura, 2001] demonstrated that $\delta-$optimality holds for the empirical likelihood ratio test.
3. In our problem,
   $\mathcal{P}_1 := \{P \in \mathcal{P} : d_L(P, P_0) \leq \lambda_1\}, \mathcal{P}_2 := \{P \in \mathcal{P} : d_L(P, P_0) \geq \lambda_2\}$.

Some definitions:

1. For any set $\Gamma \subset \mathcal{P}$, its $\delta-$smooth set is,

$$\Gamma^\delta := \cup_{\mu \in \Gamma} B(\mu, \delta).$$

2. For sets $\Gamma_1, \Gamma_2 \subseteq \mathcal{P}$, we will write

$$\inf_{\gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2} d_{KL}(\gamma_1, \gamma_2) = d_{KL}(\Gamma_1, \Gamma_2)$$

.

Theorem 1.

For any given $\delta > 0$, detector

$$\Lambda_2(n) = \Lambda_2 := \{\mu : d_{KL}(\bar{B}(\mu, 2\delta), \mathcal{P}_1) \geq \eta\}^{\delta}, \Lambda_1 := \mathcal{P}/\Lambda_2$$

is $\delta-$optimal for the deviation detection problem, i.e.,

Theorem 1.

For any given $\delta > 0$, detector

$$\Lambda_2(n) = \Lambda_2 := \{\mu : d_{KL}(\bar{B}(\mu, 2\delta), \mathcal{P}_1) \geq \eta\}^{\delta}, \Lambda_1 := \mathcal{P}/\Lambda_2$$

is $\delta-$optimal for the deviation detection problem, i.e.,

1. $\inf_{P_1 \in \mathcal{P}_1} J^{P_1}(\Lambda) \geq \eta$.

Theorem 1.

For any given $\delta > 0$, detector

$$\Lambda_2(n) = \Lambda_2 := \{\mu : d_{KL}(\bar{B}(\mu, 2\delta), \mathcal{P}_1) \geq \eta\}^{\delta}, \Lambda_1 := \mathcal{P}/\Lambda_2$$

is $\delta-$optimal for the deviation detection problem, i.e.,

1. $\inf_{P_1 \in \mathcal{P}_1} J^{P_1}(\Lambda) \geq \eta$.
2. $\inf_{P_2 \in \mathcal{P}_2} I^{P_2}(\Lambda) \geq d_{KL}(\Lambda_1, \mathcal{P}_2) := e(\eta, \delta)$.

Theorem 1.

For any given $\delta > 0$, detector

$$\Lambda_2(n) = \Lambda_2 := \{\mu : d_{KL}(\bar{B}(\mu, 2\delta), \mathcal{P}_1) \geq \eta\}^{\delta}, \Lambda_1 := \mathcal{P}/\Lambda_2$$

is $\delta-$optimal for the deviation detection problem, i.e.,

1. $\inf_{P_1 \in \mathcal{P}_1} J^{P_1}(\Lambda) \geq \eta$.
2. $\inf_{P_2 \in \mathcal{P}_2} I^{P_2}(\Lambda) \geq d_{KL}(\Lambda_1, \mathcal{P}_2) := e(\eta, \delta)$.
3. Λ is δ-optimal, i.e., if Ω is a test s.t.

$$\inf_{P_1 \in \mathcal{P}_1} J^{P_1}(\Omega^{6\delta}) \geq \eta,$$

then,

$$\inf_{P_2 \in \mathcal{P}_2} I^{P_2}(\Omega^{\delta}) \leq e(\eta, \delta).$$

Future Work

Simplify the optimal detector to be a generalized empirical likelihood ratio test ($\delta = 0$ in Theorem 1.):

$$\Lambda_2(n) = \Lambda_2 := \{\mu : d_{KL}(\mu, \mathcal{P}_1) \geq \eta\}, \qquad \Lambda_1 := \mathcal{P}/\Lambda_2.$$

Thanks