A Lower Bound for the Average Run Length in Min-Max Hypothesis Testing

David Islip

April 18, 2020

1 Background

1.1 Changepoint Detection as Hypothesis Testing

The problem of detecting a changepoint can be formulated as a hypothesis testing problem. Assuming that some sequence of observations $\{R_i\}_{i=1}^{\infty}$ (for example returns) is observed and that there are known two distributions $\mathbb{P}_1$ and $\mathbb{P}_2$, the detection of a changepoint at index κ is formulated as the following hypothesis testing problem:

$$
\begin{aligned}
H_0 &: R_i \sim \mathbb{P}_1, i = 1, 2, \dots \\
H_1 &: R_i \sim \mathbb{P}_1, i = 1, 2, \dots, \kappa, \\
&\quad R_i \sim \mathbb{P}_2, i = \kappa + 1, \kappa + 2, \dots
\end{aligned}
\tag{1}
$$

which can be constructed while controlling for multiple comparisons error from the single observation hypothesis test for a given i:

$$
H_0 : \ R_i \sim \mathbb{P}_1 \quad \text{or} \quad H_1 : \ R_i \sim \mathbb{P}_2 \tag{2}
$$

Min-max hypothesis testing allows the distributions for each hypothesis to lie within an ambiguity set, therefore, providing control over the level of robustness against distributional uncertainty.

1.2 Wasserstein Robust Hypothesis Testing

Gao et al. (2018) propose the use of order-one Wasserstein ambiguity sets for robust hypothesis testing because they allow for a suitable level of robustness via controlling the radii of the set and amount to computing the solution of convex programs that do not depend on the dimension of the data. An order-one Wasserstein ambiguity set cenetered at distribution $\mathbb{P}_n$ is defined as:

$$
\mathcal{P}_\delta(\mathbb{P}_n) := \{\mathbb{P} : D_c(\mathbb{P}, \mathbb{P}_n) \le \delta\} \tag{3}
$$

where $D_c(.)$ is the first order q norm - Wasserstein distance, defined for any two probability distributions $\mathbb{P},\ \mathbb{Q}$ supported on $\mathbb{R}^d$

$$D_c(\mathbb{P},\mathbb{Q}) := \inf\{\mathbb{E}_\pi[\|U-W\|_q] : \pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d), \pi_U = \mathbb{P}, \pi_W = \mathbb{Q}\} \tag{4}$$

where $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ denotes the space of Borel probability measures supported on $\mathbb{R}^d \times \mathbb{R}^d$ and for a given element $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$; a vector from the support of π takes the form $(U, V),\ U \in \mathbb{R}^d,\ V \in \mathbb{R}^d$. Thereby the marginals distributions are denoted as π_U and π_V.

Gao et al. (2018) then formulates the simple test (2) as the problem of finding a detector $\phi : \Omega \to \mathbb{R}$, which for a given $R \in \Omega$ accepts H_1 and rejects H_2 whenever $\phi(R) \geq 0$ and vice versa. The approach is to select a detector ϕ that minimizes the maximum of the worst case probabilities of hypothesis classification error, where the probability distributions used to evaluate the error probabilities are allowed to vary within their own order-one Wasserstein ambiguity sets centered at the assumed $\mathbb{P}_1$ and $\mathbb{P}_2$.

$$\inf_{\phi:\Omega\to\mathbb{R}} \max\big(\sup_{P_1\in\mathcal{P}^1_{\theta_1}(\mathbb{P}_1)} P_1[R : \phi(R) < 0], \sup_{P_2\in\mathcal{P}^1_{\theta_2}(\mathbb{P}_2)} P_2[R : \phi(R) \geq 0]\big) \tag{5}$$

invoking the use of a non-negative, nondecreasing convex function called the generating function l which satisfies $l(0) = 1$ and $\lim_{t\to\infty} l(t) = 0$ allows the inner supremum in (5) to be bounded by:

$$\sup_{P_1\in\mathcal{P}_{\theta_1}(\mathbb{P}_1), P_2\in\mathcal{P}_{\theta_2}(\mathbb{P}_2)} \mathbb{E}_{P_1}[l \circ (-\phi)(R)] + \mathbb{E}_{P_2}[l \circ (\phi)(R)] \tag{6}$$

because the type one and type two error are bounded their respective terms, i.e $\mathbb{P}_1(\phi(R) < 0) \leq \mathbb{E}^{\mathbb{P}_1}[l(-\phi(R))]$.

1.3 CuSum Detection Method

The cumulative sum method specifies a parameter b whereby a change is detected at the following stopping time T_1

$$T_1 = \inf\{t > 0 | \max_{1\leq k\leq t} \sum_{i=k}^{t} (-\phi(R^i)) \geq b\} \tag{7}$$

2 Average Run Length for any Loss function

Proposition: If (5) has a feasible solution with objective ϵ^{**} less than 0.5, the ARL of the resulting detector using the CuSum Method can be bounded below by the following inequality:

$$\frac{1}{ARL} \leq \sum_{k=1}^{\infty} \Big(\mathbb{E}^{\mathbb{P}_1}[l(-\phi^*_k(R))]\Big)^k \tag{8}$$

Where $\phi^*_k = \phi + \frac{b}{k}$

Proof: First, recall:

$$\mathbb{P}_1(\phi(R) < 0) \le \mathbb{E}^{\mathbb{P}_1}[l(-\phi(R))] \tag{9}$$

The formula above is only for the simple hypothesis test therefore to evaluate/get a bound for the type one error for the CUSUM type detector one can follow the logic from appendix A of (Cao & Xie, 2017). The proof follows:

First note: $T = \inf\{t > 0 | \sum_{i=1}^{t} -\phi(R^i) > b\}$ is the same procedure as T_1 and is always larger than T_1

$$\begin{aligned}
\mathbb{P}_1(T \le m) &\le \mathbb{P}_1\Big(\bigcup_{k=1}^{m}\Big\{\sum_{i=1}^{k} -\phi(R^i) > b\Big\}\Big) && (10)\\
&\le \sum_{k=1}^{m} \mathbb{P}_1\Big(\sum_{i=1}^{k} -\phi(R^i) > b\Big) && (11)\\
&= \sum_{k=1}^{m} \mathbb{P}_1\Big(\sum_{i=1}^{k}(-\phi(R^i) - \frac{b}{k}) > 0\Big) && (12)
\end{aligned}$$

Letting $\phi_k^* = \phi + \frac{b}{k}$, implies:

$$\mathbb{P}_1(-\phi_k^*(R) > 0) \le \mathbb{E}^{\mathbb{P}_1}[l(-\phi_k^*(R))]$$

Therefore in the case of k repeated observations:

$$\mathbb{P}_1\Big(\sum_{i=1}^{k}(-\phi(R^i) - \frac{b}{k}) > 0\Big) \le \mathbb{E}^{\mathbb{P}_1}[l(-\phi_k^*(R))]^k$$

Then it follows that

$$\mathbb{P}_1(T \le m) \le \sum_{k=1}^{m}\Big(\mathbb{E}^{\mathbb{P}_1}[l(-\phi_k^*(R))]\Big)^k \tag{13}$$

The infinite sum is finite, because by Theorem 1 in Gao et al. (2018), $\epsilon^* * < 0.5$ implies $\mathbb{E}^{\mathbb{P}_1}[l(-\phi(R))] < 1$ and therefore $\mathbb{E}^{\mathbb{P}_1}[l(-\phi_k^*(R))]$ since $\phi_k > \phi$ and l is non negative and non decreasing. Lastly, since $\mathbb{E}^{\mathbb{P}_1}[T] \ge \frac{1}{\mathbb{P}_1(T<\infty)}$ the result follows ∎

Note that expressions depend on the form of the generating function. For example Goldenshluger et al use $\phi = \log(p_1/p_2)$ as a detector with $l(.) = \exp(.)$ and find that $\mathbb{E}^{\mathbb{P}_1}[\exp(-\phi^*(\omega))] \le \exp(\frac{-b}{k})\epsilon *$ which allows Cao et. al to derive that $\mathbb{P}_{\not{1}}(T < \infty) = \exp(-b) * \frac{\epsilon^*}{(1-\epsilon^*)}$

2.1 Computation

Computing the sum is tractable because at any step one can cheaply calculate an upper bound on the remaining terms. using a geometric approximation

$$\sum_{k=1}^{\infty}\left(\mathbb{E}^{\mathbb{P}_{\nVdash}}[l(1-\phi(R)-\frac{b}{k})]\right)^{k} \leq \sum_{k=1}^{M}\left(\mathbb{E}^{\mathbb{P}_{\nVdash}}[l(1-\phi(R)-\frac{b}{k})]\right)^{k} + \sum_{k=M+1}^{\infty}\left(\mathbb{E}^{\mathbb{P}_{\nVdash}}[l(1-\phi(R))]\right)^{k} \tag{14}$$

At each step M we know that there is less than

$$\sum_{k=M+1}^{\infty}\left(\mathbb{E}^{\mathbb{P}_{\nVdash}}[l(1-\phi(R))]\right)^{k} = \frac{\mathbb{E}^{\mathbb{P}_{\nVdash}}[l(1-\phi(R))]^{M}}{(1-\mathbb{E}^{\mathbb{P}_{\nVdash}}[l(1-\phi(R))])} \tag{15}$$

remaining in the sum.