

# Should I Stop or Should I Go: Updated Theorems and Proofs

## 1. Theorems

**Theorem 3.2** (CLASH weights converge to optimal weights). *The CLASH weights satisfy the error bound*

$$|\hat{w}_i^n - w_i^*| \leq \exp \left\{ -\frac{(\tau(x_i) - \delta)^2}{2\hat{\sigma}_{n,-i}^2(x_i)} \right\} + \frac{|\hat{\tau}_{n,-i}(x_i) - \tau(x_i)|}{\hat{\sigma}_{n,-i}(x_i)} \exp \left\{ -\frac{(\tau(x_i) - \delta)^2}{2\hat{\sigma}_{n,-i}^2(x_i)} + \frac{|\tau(x_i) - \delta| |\hat{\tau}_{n,-i}(x_i) - \tau(x_i)|}{\hat{\sigma}_{n,-i}^2(x_i)} \right\}.$$

Moreover, if  $\delta < \inf_{x:\tau(x)>0} \tau(x)$  and, given  $x_i$ ,  $\hat{\tau}_{n,-i}(x_i) \xrightarrow{P} \tau(x_i)$  and  $\hat{\sigma}_{n,-i}(x_i) \xrightarrow{P} 0$ , then  $\hat{w}_i^n$  is a consistent estimator of the optimal weight:  $\hat{w}_i^n - w_i^* \xrightarrow{P} 0$ .

**Theorem 3.3** (CLASH limits unnecessary stopping). *Consider a stopping test with weighted z-statistic and weights estimated using CLASH. If  $\max_{i \leq n} \hat{\sigma}_{n,-i}^2(x_i) = o_p(1/\log(n))$ ,  $\max_{i \leq n} |\tau(x_i) - \hat{\tau}_{n,-i}(x_i)| = o_p(1)$ , and  $y_i$  are uniformly bounded, then the stopping probability of the test converges to zero if no participant group is harmed.*

## 2. Proofs

We now present proofs for Thm. 3.2 and Thm. 3.3. Note that we repeatedly use the following property. For any events  $A$  and  $B$ ,

$$\mathbb{P}(A, B) \leq \min\{\mathbb{P}(A), \mathbb{P}(B)\}. \quad (5)$$

### 2.1. Proof of Thm. 3.2

Our entire proof will be carried out conditional on the value  $x_i$ . Recall that we define our weights as

$$\hat{w}_i^n = 1 - \Phi\left(\frac{\delta - \hat{\tau}_{n,-i}(x_i)}{\hat{\sigma}_{n,-i}(x_i)}\right).$$

**Error Bound** We first establish a bound on the difference between our estimated weights  $\hat{w}_i^n$  and the optimal weights  $w_i^*$ . Consider two cases.

**Case 1:**  $\tau(x_i) \leq 0$ . In this case,  $w_i^* = 0$  by definition, and so we just need to prove a bound on the magnitude of  $\hat{w}_i^n$ . Using Taylor's theorem with Lagrange remainder, we know that  $\exists h_n \in [0, 1]$  such that

$$\begin{aligned} \hat{w}_i^n &= 1 - \Phi\left(\frac{\delta - \hat{\tau}_{n,-i}(x_i)}{\hat{\sigma}_{n,-i}(x_i)}\right) \\ &= 1 - \Phi\left(\frac{\delta - \tau(x_i)}{\hat{\sigma}_{n,-i}(x_i)}\right) - \frac{\tau(x_i) - \hat{\tau}_{n,-i}(x_i)}{\hat{\sigma}_{n,-i}(x_i)} \phi\left(\frac{\delta - \tau(x_i)}{\hat{\sigma}_{n,-i}(x_i)} + h_n \frac{\tau(x_i) - \hat{\tau}_{n,-i}(x_i)}{\hat{\sigma}_{n,-i}(x_i)}\right) \\ &\leq 1 - \Phi\left(\frac{\delta - \tau(x_i)}{\hat{\sigma}_{n,-i}(x_i)}\right) + \frac{|\tau(x_i) - \hat{\tau}_{n,-i}(x_i)|}{\hat{\sigma}_{n,-i}(x_i)} \phi\left(\frac{\delta - \tau(x_i)}{\hat{\sigma}_{n,-i}(x_i)} + h_n \frac{\tau(x_i) - \hat{\tau}_{n,-i}(x_i)}{\hat{\sigma}_{n,-i}(x_i)}\right) \end{aligned}$$

where  $\phi$  is the standard Gaussian probability density function. Since  $\delta - \tau(x_i) > 0$ , we can use the Chernoff inequality to bound the first term,

$$1 - \Phi\left(\frac{\delta - \tau(x_i)}{\hat{\sigma}_{n,-i}(x_i)}\right) \leq \exp\left(-\frac{(\delta - \tau(x_i))^2}{2\hat{\sigma}_{n,-i}^2(x_i)}\right).$$

We now focus on the second term.

$$\begin{aligned}
 \phi\left(\frac{\delta - \tau(x_i)}{\hat{\sigma}_{n,-i}(x_i)} + h_n \frac{\tau(x_i) - \hat{\tau}_{n,-i}(x_i)}{\hat{\sigma}_{n,-i}(x_i)}\right) &= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\delta - \tau(x_i)}{\hat{\sigma}_{n,-i}(x_i)} + h_n \frac{\tau(x_i) - \hat{\tau}_{n,-i}(x_i)}{\hat{\sigma}_{n,-i}(x_i)}\right)^2\right\} \\
 &= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\left(\frac{\delta - \tau(x_i)}{\hat{\sigma}_{n,-i}(x_i)}\right)^2 + h_n^2 \left(\frac{\tau(x_i) - \hat{\tau}_{n,-i}(x_i)}{\hat{\sigma}_{n,-i}(x_i)}\right)^2 - \right.\right. \\
 &\quad \left.\left.2h_n \frac{(\delta - \tau(x_i))(\hat{\tau}_{n,-i}(x_i) - \tau(x_i))}{\hat{\sigma}_{n,-i}^2(x_i)}\right]\right\} \\
 &\leq \exp\left\{-\frac{1}{2}\left[\left(\frac{\delta - \tau(x_i)}{\hat{\sigma}_{n,-i}(x_i)}\right)^2 - 2h_n \frac{(\delta - \tau(x_i))|\hat{\tau}_{n,-i}(x_i) - \tau(x_i)|}{\hat{\sigma}_{n,-i}^2(x_i)}\right]\right\} \\
 &\leq \exp\left\{-\frac{1}{2}\left[\left(\frac{\delta - \tau(x_i)}{\hat{\sigma}_{n,-i}(x_i)}\right)^2 - 2 \frac{(\delta - \tau(x_i))|\hat{\tau}_{n,-i}(x_i) - \tau(x_i)|}{\hat{\sigma}_{n,-i}^2(x_i)}\right]\right\},
 \end{aligned}$$

since  $h_n \in [0, 1]$ . Thus, we have that

$$\hat{w}_i^n \leq \exp\left\{-\frac{(\delta - \tau(x_i))^2}{2\hat{\sigma}_{n,-i}^2(x_i)}\right\} + \frac{|\tau(x_i) - \hat{\tau}_{n,-i}(x_i)|}{\hat{\sigma}_{n,-i}(x_i)} \exp\left\{-\frac{(\delta - \tau(x_i))^2}{2\hat{\sigma}_{n,-i}^2(x_i)} + \frac{(\delta - \tau(x_i))|\hat{\tau}_{n,-i}(x_i) - \tau(x_i)|}{\hat{\sigma}_{n,-i}^2(x_i)}\right\}.$$

**Case 2:**  $\tau(x_i) > 0$ . In this case,  $w_i^* = 1$ . Thus,

$$\begin{aligned}
 |\hat{w}_i^n - w_i^*| &= 1 - \hat{w}_i^n \\
 &= \Phi\left(\frac{\delta - \hat{\tau}_{n,-i}(x_i)}{\hat{\sigma}_i(x_i)}\right) \\
 &= \Phi\left(\frac{\delta - \tau(x_i)}{\hat{\sigma}_{n,-i}(x_i)}\right) + \frac{|\tau(x_i) - \hat{\tau}_{n,-i}(x_i)|}{\hat{\sigma}_{n,-i}(x_i)} \phi\left(\frac{\delta - \tau(x_i)}{\hat{\sigma}_{n,-i}(x_i)} + h_n \frac{\tau(x_i) - \hat{\tau}_{n,-i}(x_i)}{\hat{\sigma}_{n,-i}(x_i)}\right) \\
 &= 1 - \Phi\left(\frac{\tau(x_i) - \delta}{\hat{\sigma}_{n,-i}(x_i)}\right) + \frac{|\tau(x_i) - \hat{\tau}_{n,-i}(x_i)|}{\hat{\sigma}_{n,-i}(x_i)} \phi\left(\frac{\delta - \tau(x_i)}{\hat{\sigma}_{n,-i}(x_i)} + h_n \frac{\tau(x_i) - \hat{\tau}_{n,-i}(x_i)}{\hat{\sigma}_{n,-i}(x_i)}\right) \\
 &\leq \exp\left\{-\frac{(\delta - \tau(x_i))^2}{2\hat{\sigma}_{n,-i}^2(x_i)}\right\} + \\
 &\quad \frac{|\tau(x_i) - \hat{\tau}_{n,-i}(x_i)|}{\hat{\sigma}_{n,-i}(x_i)} \exp\left\{-\frac{(\delta - \tau(x_i))^2}{2\hat{\sigma}_{n,-i}^2(x_i)} + \frac{(\tau(x_i) - \delta)|\hat{\tau}_{n,-i}(x_i) - \tau(x_i)|}{\hat{\sigma}_{n,-i}^2(x_i)}\right\}.
 \end{aligned}$$

using the same Taylor expansion and Chernoff bound from above.

In summary, we have established that

$$|\hat{w}_i^n - w_i^*| \leq \exp\left\{-\frac{(\delta - \tau(x_i))^2}{2\hat{\sigma}_{n,-i}^2(x_i)}\right\} + \frac{|\tau(x_i) - \hat{\tau}_{n,-i}(x_i)|}{\hat{\sigma}_{n,-i}(x_i)} \exp\left\{-\frac{(\delta - \tau(x_i))^2}{2\hat{\sigma}_{n,-i}^2(x_i)} + \frac{|\tau(x_i) - \delta||\hat{\tau}_{n,-i}(x_i) - \tau(x_i)|}{\hat{\sigma}_{n,-i}^2(x_i)}\right\}.$$

**Consistency of  $\hat{w}_i^n$**  We now establish the consistency of  $\hat{w}_i^n$  using the derived bound. We assume that  $\delta < \inf_{x: \tau(x) > 0} \tau(x)$  and, given  $x_i$ ,  $\hat{\tau}_{n,-i}(x_i) \xrightarrow{P} \tau(x_i)$  and  $\hat{\sigma}_{n,-i}(x_i) \xrightarrow{P} 0$ .

Define  $a_i = |\tau(x_i) - \delta|/\sqrt{2}$  and  $Z_{i,n} = \frac{\tau(x_i) - \hat{\tau}_{n,-i}(x_i)}{\hat{\sigma}_{n,-i}(x_i)}$ . Note that  $a_i$  is strictly positive and a constant (given  $x_i$ ). From the error bound, we have that

$$|\hat{w}_i^n - w_i^*| \leq \exp(-a_i^2/\hat{\sigma}_{n,-i}(x_i)^2) + |Z_{i,n}| \exp[-a_i^2/\hat{\sigma}_{n,-i}(x_i)^2 + \sqrt{2}a_i|Z_{i,n}|/\hat{\sigma}_{n,-i}(x_i)].$$

Now, we fix any  $\epsilon > 0$ . Applying the law of total probability and equation (5), we find that

$$\begin{aligned}
 \mathbb{P}(|\hat{w}_i^n - w_i^*| > \epsilon) &\leq \mathbb{P}(\exp(-a_i^2/\hat{\sigma}_{n,-i}(x_i)^2) + |Z_{i,n}| \exp[-a_i^2/\hat{\sigma}_{n,-i}(x_i)^2 + \sqrt{2}a_i|Z_{i,n}|/\hat{\sigma}_{n,-i}(x_i)] > \epsilon) \\
 &= \mathbb{P}(\exp(-a_i^2/\hat{\sigma}_{n,-i}(x_i)^2) + |Z_{i,n}| \exp[-a_i^2/\hat{\sigma}_{n,-i}(x_i)^2 + \sqrt{2}a_i|Z_{i,n}|/\hat{\sigma}_{n,-i}(x_i)] > \epsilon, \\
 &\quad |Z_{i,n}\hat{\sigma}_{n,-i}(x_i)| > a_i/(2\sqrt{2})) + \\
 &\quad \mathbb{P}(\exp(-a_i^2/\hat{\sigma}_{n,-i}(x_i)^2) + |Z_{i,n}| \exp[-a_i^2/\hat{\sigma}_{n,-i}(x_i)^2 + \sqrt{2}a_i|Z_{i,n}|/\hat{\sigma}_{n,-i}(x_i)] > \epsilon, \\
 &\quad |Z_{i,n}\hat{\sigma}_{n,-i}(x_i)| \leq a_i/(2\sqrt{2})) \\
 &\leq \mathbb{P}(|Z_{i,n}\hat{\sigma}_{n,-i}(x_i)| > a_i/(2\sqrt{2})) + \\
 &\quad \mathbb{P}(\exp(-a_i^2/\hat{\sigma}_{n,-i}(x_i)^2) + |Z_{i,n}| \exp[-a_i^2/\hat{\sigma}_{n,-i}(x_i)^2 + \sqrt{2}a_i|Z_{i,n}|/\hat{\sigma}_{n,-i}(x_i)] > \epsilon, \\
 &\quad |Z_{i,n}\hat{\sigma}_{n,-i}(x_i)| \leq a_i/(2\sqrt{2})) \\
 &\leq \mathbb{P}(|Z_{i,n}\hat{\sigma}_{n,-i}(x_i)| > a_i/(2\sqrt{2})) + \\
 &\quad \mathbb{P}\left(\left(1 + \frac{a_i}{2\sqrt{2}\hat{\sigma}_{n,-i}(x_i)}\right) \exp\left(\frac{-a_i^2}{2\hat{\sigma}_{n,-i}(x_i)^2}\right) > \epsilon, |Z_{i,n}\hat{\sigma}_{n,-i}(x_i)| \leq a_i/(2\sqrt{2})\right) \\
 &\leq \mathbb{P}(|Z_{i,n}\hat{\sigma}_{n,-i}(x_i)| > a_i/(2\sqrt{2})) + \mathbb{P}\left(\left(1 + \frac{a_i}{2\sqrt{2}\hat{\sigma}_{n,-i}(x_i)}\right) \exp\left(\frac{-a_i^2}{2\hat{\sigma}_{n,-i}(x_i)^2}\right) > \epsilon\right)
 \end{aligned}$$

where the second last inequality follows by substituting  $|Z_{i,n}\hat{\sigma}_{n,-i}(x_i)| \leq a_i/(2\sqrt{2})$  into the bound and algebraically simplifying and the last inequality follows from (5). Now, the first term on the right converges to 0 since  $Z_{i,n}\hat{\sigma}_{n,-i}(x_i) = \tau(x_i) - \hat{\tau}_{n,-i}(x_i) \xrightarrow{P} 0$ . We examine the second term. Define  $\xi_{i,n} = a_i/(\sqrt{2}\hat{\sigma}_{n,-i}(x_i))$ . Since a convex function is no smaller than its tangent line, we have  $(1 + \xi_{i,n}/2) \leq \exp(\xi_{i,n}/2)$ . Then,

$$\begin{aligned}
 \mathbb{P}\left(\left(1 + \frac{a_i}{2\sqrt{2}\hat{\sigma}_{n,-i}(x_i)}\right) \exp\left(\frac{-a_i^2}{2\hat{\sigma}_{n,-i}(x_i)^2}\right) > \epsilon\right) &\leq \mathbb{P}((1 + \xi_{i,n}/2) \exp(-\xi_{i,n}^2) > \epsilon) \\
 &\leq \mathbb{P}(\exp(-\xi_{i,n}^2 + \xi_{i,n}/2) > \epsilon)
 \end{aligned}$$

which converges to 0 as  $\hat{\sigma}_{n,-i}(x_i) \xrightarrow{P} 0$  (and thus  $\xi_{i,n}$  diverges in probability to  $\infty$ ). Thus, we have shown that

$$\mathbb{P}(|\hat{w}_i^n - w_i^*| > \epsilon) \rightarrow 0,$$

which establishes the desired consistency.

## 2.2. Proof of Thm. 3.3

Suppose that no participant group is harmed so that  $w_i^* = 0$  for all  $i$ . Let  $a = \delta/\sqrt{2}$ , and  $a_i = |\tau(x_i) - \delta|/\sqrt{2}$ , and note that  $a \leq a_i$  for each  $i$  as each  $\tau(x_i) \leq 0$ . Thus, if  $|Z_{i,n}\hat{\sigma}_{n,-i}(x_i)| \leq a/(2\sqrt{2})$ , then  $|Z_{i,n}\hat{\sigma}_{n,-i}(x_i)| \leq a_i/(2\sqrt{2})$  and hence, as in the proof of Thm. 3.2,

$$\hat{w}_i^n \leq \left(1 + \frac{a_i}{2\sqrt{2}\hat{\sigma}_{n,-i}(x_i)}\right) \exp\left(\frac{-a_i^2}{2\hat{\sigma}_{n,-i}(x_i)^2}\right)$$

Fix any  $\epsilon > 0$ , let  $\xi_{i,n} = a_i/(\sqrt{2}\hat{\sigma}_{n,-i}(x_i))$ , and define  $f(x) = (1 + x/2) \exp(-x^2)$ . As in the proof of Thm. 3.2, we can apply the law of total probability and equation (5) to write,

$$\begin{aligned}
 \mathbb{P}(\sum_i \hat{w}_i^n > \epsilon) &\leq \mathbb{P}(\max_{i \leq n} |Z_{i,n}\hat{\sigma}_{n,-i}(x_i)| > a/(2\sqrt{2})) + \mathbb{P}(\sum_i \hat{w}_i^n > \epsilon, \max_{i \leq n} |Z_{i,n}\hat{\sigma}_{n,-i}(x_i)| \leq a/(2\sqrt{2})) \\
 &\leq \mathbb{P}(\max_{i \leq n} |Z_{i,n}\hat{\sigma}_{n,-i}(x_i)| > a/(2\sqrt{2})) + \\
 &\quad \mathbb{P}(\sum_i (1 + a_i/(2\sqrt{2}\hat{\sigma}_{n,-i}(x_i))) \exp(-a_i^2/(2\hat{\sigma}_{n,-i}(x_i)^2)) > \epsilon) \\
 &= \mathbb{P}(\max_{i \leq n} |Z_{i,n}\hat{\sigma}_{n,-i}(x_i)| > a/(2\sqrt{2})) + \mathbb{P}(\sum_i f(\xi_{i,n}) > \epsilon).
 \end{aligned}$$

Since  $\max_{i \leq n} |Z_{i,n} \hat{\sigma}_{n,-i}(x_i)| = o_p(1)$  by assumption, we have  $\mathbb{P}(\max_{i \leq n} |Z_{i,n} \hat{\sigma}_{n,-i}(x_i)| > a/(2\sqrt{2})) \rightarrow 0$ . Meanwhile, since a convex function is no smaller than its tangent line, we have  $(1 + x/2) \leq \exp(x/2)$  and hence

$$f(x) \leq \exp(-x^2 + x/2) \leq \exp(-x^2 + x^2/2 + 1/8) = \exp(-x^2/2 + 1/8)$$

where we used the arithmetic-geometric mean inequality in the final inequality. Therefore, since  $a \leq \min_i a_i$ ,

$$\begin{aligned} \mathbb{P}(\sum_i f(\xi_{i,n}) > \epsilon) &\leq \mathbb{P}(\sum_i \exp(-\xi_{i,n}^2/2 + 1/8) > \epsilon) \leq \mathbb{P}(n \exp(-(\min_{i \leq n} \xi_{i,n}^2)/2 + 1/8) > \epsilon) \\ &\leq \mathbb{P}(n \exp(-(\min_{i \leq n} a_i^2/\hat{\sigma}_{n,-i}^2(x_i))/4 + 1/8) > \epsilon) \\ &\leq \mathbb{P}(\exp(-a^2/(4 \max_{i \leq n} \hat{\sigma}_{n,-i}^2(x_i)) + 1/8 + \log n) > \epsilon) \end{aligned}$$

Since  $\max_{i \leq n} \hat{\sigma}_{n,-i}^2(x_i) = o_p(1/\log(n))$  by assumption, we further have  $\mathbb{P}(\sum_i f(\xi_{i,n}) > \epsilon) \rightarrow 0$ . Since  $\epsilon > 0$  was arbitrary, we have shown that  $\sum_{i=1}^n \hat{w}_i^n \xrightarrow{p} 0$ .

Now, recall the form of the CLASH weighted z-statistic

$$\lambda_n^w = \frac{\sqrt{\sum_{i=1}^n \hat{w}_i^n}}{\sqrt{2\sigma^2}} \left( \frac{\sum_{i=1}^n \hat{w}_i^n y_i d_i}{\sum_{i=1}^n \hat{w}_i^n d_i} - \frac{\sum_{i=1}^n \hat{w}_i^n y_i (1 - d_i)}{\sum_{i=1}^n \hat{w}_i^n (1 - d_i)} \right)$$

Define  $c$  such that  $|y_i| \leq c$ . We know  $c$  must exist, since the outcomes are bounded by assumption. Then, we have that

$$\begin{aligned} |\lambda_n^w| &= \left| \frac{\sqrt{\sum_{i=1}^n \hat{w}_i^n}}{\sqrt{2\sigma^2}} \left( \frac{\sum_{i=1}^n \hat{w}_i^n y_i d_i}{\sum_{i=1}^n \hat{w}_i^n d_i} - \frac{\sum_{i=1}^n \hat{w}_i^n y_i (1 - d_i)}{\sum_{i=1}^n \hat{w}_i^n (1 - d_i)} \right) \right| \\ &\leq \frac{\sqrt{\sum_{i=1}^n \hat{w}_i^n}}{\sqrt{2\sigma^2}} \left( \frac{\sum_{i=1}^n \hat{w}_i^n c d_i}{\sum_{i=1}^n \hat{w}_i^n d_i} - \frac{\sum_{i=1}^n \hat{w}_i^n (-c)(1 - d_i)}{\sum_{i=1}^n \hat{w}_i^n (1 - d_i)} \right) \\ &= \frac{\sqrt{\sum_{i=1}^n \hat{w}_i^n}}{\sqrt{2\sigma^2}} 2c \\ &\xrightarrow{p} 0 \end{aligned}$$

Thus, we see that the weighted test statistic  $\lambda_n^w$  converges in probability to 0. Now, the test can only reject if  $\lambda_n^w$  exceeds a fixed and positive bound  $b_\alpha$ . By the definition of convergence in probability, this probability must shrink to zero.