Optimal Stopping and Worker Selection in Crowdsourcing: an Adaptive Sequential Probability Ratio Test Framework

Xiaoou Li[1], *Yunxiao Chen*[2], *Xi Chen*[3], *Jingchen Liu*[4], *and Zhiliang Ying*[4]

University of Minnesota[1], *London School of Economics and Political Science*[2], *New York University*[3], *and Columbia University*[4]

Supplementary Material

In the supplement material, we present the proof of Proposition 1, and Theorem 1, 2 and 3. The proof for supporting lemmas are presented in Section S2. We also present simulated experiments in Section S3.

S1 Proofs of Technical Results

S1.1 Proof of Proposition 1

We consider the more general problem of finding the optimal future sequential adaptive design after collecting n samples. Suppose that the first n responses are $x_1, ..., x_n$ and the first n experiment selection functions are $j_1, ..., j_n$. We need to decide the experiment selection function for the $(n+1)$'s sample, that is, $j_{n+1}(x_1, ..., x_n)$. We also need to decide whether to stop the test or not and if the test is stopped, which hypothesis should be chosen. We first consider the stopping rule. To describe the stopping rule, we define the loss function

$$L\{(N, D), \theta\} = \mathbf{1}_{\{D\neq\theta\}} + cN, \tag{S1.1}$$

and the conditional risk for a test procedure (J, N, D) of stopping the test with n samples,

$$\mathbb{E}\Big[L\{(N, D), \theta\}\Big| X_{1:n} = x_{1:n}, N = n\Big], \tag{S1.2}$$

where we write $x_{1:n}$ as the abbreviation for the sequence $(x_1, ..., x_n)$. Because $\mathbb{E}\mathbf{1}_{\{D\neq\theta\}} = \mathbb{P}(\theta = 0|X_{1:n} = x_{1:n})\mathbf{1}_{\{D=1\}} + \mathbb{P}(\theta = 1|X_{1:n} = x_{1:n})\mathbf{1}_{\{D=0\}}$, it is straightforward that given $N = n$ and $X_{1:n} = x_{1:n}$, the optimal decision D is

$$D = 1 \text{ if } \mathbb{P}(\theta = 1|X_{1:n} = x_{1:n}) \geq \mathbb{P}(\theta = 0|X_{1:n} = x_{1:n}) \text{ and } D = 0 \text{ otherwise.} \tag{S1.3}$$

We insert this to (S1.2) and obtain the minimal conditional risk for stopping the test with n samples,

$$\begin{aligned} r_s(x_{1:n}, j_{1:n}) &= \inf_D \mathbb{E}\Big[L\{(N, D), \theta\}\Big| X_{1:n} = x_{1:n}, N = n\Big] \qquad \text{(S1.4)} \\ &= \min\{\mathbb{P}(\theta = 0|X_{1:n} = x_{1:n}), \mathbb{P}(\theta = 1|X_{1:n} = x_{1:n})\} + nc. \end{aligned}$$

We proceed to the minimal conditional risk for continuing the test with at least $n+1$ samples,

$$r_c(x_{1:n}, j_{1:n}) = \inf_{(J,N,D)\in\mathcal{A}_{x_{1:n},j_{1:n}}} \mathbb{E}\Big[L\{(N, D), \theta\}\Big| X_{1:n} = x_{1:n}\Big], \tag{S1.5}$$

where the set $\mathcal{A}_{x_{1:n},j_{1:n}}$ consists of all the sequential adaptive designs that have $j_{1:n}$ as the first n experiment selection function and do not stop with $x_{1:n}$ as the first n observations.

Clearly, the optimal test should continue to collect more samples if the minimal conditional risk for continuing the test is smaller than the minimal conditional risk for stopping the test. That is, the test is stopped if and only if

$$g(x_{1:n}, j_{1:n}) \leq 0,$$

where g is the maximal reduced conditional risk,

$$g(x_{1:n}, j_{1:n}) = r_s(x_{1:n}, j_{1:n}) - r_c(x_{1:n}, j_{1:n}). \tag{S1.6}$$

The function $g(x_{1:n}, j_{1:n})$ determines a continuing region $\{(X_1, ..., X_n) : g(X_{1:n}, j_{1:n}) > 0\}$ for the sequence of samples. We further explore the shape of the continuing region. We abuse the notation a little and define the log-likelihood function

$$l(x_{1:n}, j_{1:n}) = \log\Big(\frac{\prod_{i=1}^{n} f_{1,\delta_i}(x_i)}{\prod_{i=1}^{n} f_{0,\delta_i}(x_i)}\Big), \tag{S1.7}$$

where $\delta_i = j_i(x_{1:i-1})$ is the i-th selected experiment for $i = 1, ...n$. The following lemma, whose proof is provided in Section S2, shows that the function g depends only on the log-likelihood ratio.

Lemma 1. *There exists a function $h : \mathbb{R} \to \mathbb{R}$ such that for all sequence of observations $x_{1:n}$ and experiment selection functions $j_{1:n}$,*

$$g(x_{1:n}, j_{1:n}) = h(l(x_{1:n}, j_{1:n})).$$

According to Lemma 1 and the previous analysis, the optimal stopping rule is determined through the continuing region of the likelihood ratio. That is, the stopping time for the optimal design is

$$N^* = \inf\{n : l(X_{1:n}, j^*_{1:n}) \notin C\},$$

where

$$C = h^{-1}(0, \infty). \tag{S1.8}$$

and $j^*_{1:n}$ is the sequence of experiment selection functions for the optimal design. Furthermore, we describe the shape of the continuing region C in the following lemma, whose proof is given in Section S2.

Lemma 2. *If $a > b > \log \frac{\pi_0}{\pi_1}$ and $a \in C$, then $b \in C$. Similarly, if $a < b < \log \frac{\pi_0}{\pi_1}$ and $a \in C$, then $b \in C$.*

Lemma 2 implies that the continuing region is an interval that $C = (B, A)$ for some boundary values A and B. This completes our proof for Proposition 1(ii). In addition, we have

$$\mathbb{P}(\theta = 0 | X_1, ..., X_n) = \frac{\pi_0}{\pi_0 + \pi_1 e^{l_n}} \text{ and } \mathbb{P}(\theta = 1 | X_1, ..., X_n) = \frac{\pi_1 e^{l_n}}{\pi_0 + \pi_1 e^{l_n}}. \tag{S1.9}$$

We insert this to (S1.3) and Proposition 1(iii) is proved.

For the rest of the proof, we consider the optimal experiment selection. Considering the best choice between stopping the test and continuing the test, the minimal conditional risk given the first n samples $x_{1:n}$ is defined as

$$U_n(x_{1:n}, j_{1:n}) = \min\{r_s(x_{1:n}, j_{1:n}), r_c(x_{1:n}, j_{1:n})\}. \tag{S1.10}$$

The optimal $(n+1)$-th experiment selection $j_{n+1}(x_{1:n})$ minimizes the future conditional risk

$$j_{n+1}(x_{1:n}) = \arg \inf_{j_{n+1}(x_{1:n})} \mathbb{E}\Big[U_{n+1}(X_{1:n+1}, j_{1:n+1}) \Big| X_{1:n} = x_{1:n}\Big]. \tag{S1.11}$$

Just a clarification that if the test is stopped with the first n samples, then the choice of $j_{n+1}(x_{1:n})$ and does not affect the conditional risk and is thus arbitrary. We simplify the conditional expectation in the above display

$$
\begin{aligned}
U_{n+1}(X_{1:n+1}, j_{1:n+1}) =& \min\{r_c(X_{1:n+1}, j_{1:n+1}), r_s(X_{1:n+1}, j_{1:n+1})\}\\
=& r_s(X_{1:n+1}, j_{1:n+1}) - g(X_{1:n+1}, j_{1:n+1})_+,
\end{aligned}
$$

where the function g is defined in (S1.6) and $x_+ = \max(x, 0)$. According to Lemma 1 and (S1.4), we have

$$
\mathbb{E}\Big[U_{n+1}(X_{1:n+1}, j_{1:n+1})\Big|X_{1:n} = x_{1:n}\Big] = (n+1)c + \mathbb{E}\Big[u(l_{n+1})|X_{1:n} = x_{1:n}\Big], \tag{S1.12}
$$

where the function u is defined as

$$
u(l) = \min\{\frac{\pi_0}{\pi_0 + \pi_1 e^l}, \frac{\pi_1 e^l}{\pi_0 + \pi_1 e^l}\} - h(l)_+,
$$

and $h(l)$ is defined in Lemma 1. Consequently, (S1.11) can be written as

$$
\begin{aligned}
& j_{n+1}(x_{1:n}) \\
= \quad & \arg\inf_{j_{n+1}(x_{1:n})} \Big\{ \quad \mathbb{P}(\theta = 0|X_{1:n} = x_{1:n})\mathbb{E}[u(l_{n+1})|X_{1:n} = x_{1:n}, \theta = 0] \\
& \qquad +\mathbb{P}(\theta = 1|X_{1:n} = x_{1:n})\mathbb{E}[u(l_{n+1})|X_{1:n} = x_{1:n}, \theta = 1]\Big\}. \qquad \text{(S1.13)}
\end{aligned}
$$

Notice that $l_{n+1} = l_n + \log \frac{f_{1,j_{n+1}(x_{1:n})}(X_{n+1})}{f_{0,j_{n+1}(x_{1:n})}(X_{n+1})}$ and posterior of θ is given in (S1.9). Therefore, (S1.13) can be written as

$$
j_{n+1}(x_{1:n}) = \arg \inf_{j_{n+1}(x_{1:n})} v\Big(l_n, j_{n+1}(x_{1:n})\Big)
$$

for some bivariate function v. Let the function $j^*(l) = \arg\inf_\delta v(l, \delta)$. Then, we have $j_{n+1}(x_{1:n}) = j^*(l_n)$, and Proposition 1(i) is proved.

S1.2 Proof of Theorem 1

Similar to the proof of Proposition 1, the stopping rule for the truncated test is determined by the maximal reduced conditional risk function

$$g^{\dagger}(x_{1:n}, j_{1:n}) = r_s(x_{1:n}, j_{1:n}) - r^{\dagger}_{nc}(x_{1:n}, j_{1:n}),$$

where r_s is defined in (S1.4), and $r^{\dagger}_{nc}$ is defined similarly to (S1.5),

$$r^{\dagger}_{nc} = \inf_{(J,N,D)\in\mathcal{A}^T_{x_{1:n},\delta_{1:n}}} \mathbb{E}\Big[L\{(N,D),\theta\}\Big|X_{1:n} = x_{1:n}\Big]$$

and $\mathcal{A}^T_{x_{1:n},j_{1:n}}$ consists of all sequential adaptive design that belongs to $\mathcal{A}_{x_{1:n},j_{1:n}}$ and has a truncation length T. Similar to Lemma 1, we establish the following lemma, whose proof is similar to the proof of Lemma 1 and that of Lemma 2.

Lemma 3. *There exists a function $h^{\dagger} : \mathbb{R} \times \mathbb{Z}_+ \to \mathbb{R}$ such that*

$$g^{\dagger}(x_{1:n}, j_{1:n}) = h^{\dagger}(l(x_{1:n}, j_{1:n}), n). \qquad \text{(S1.14)}$$

In addition, for $n = 1, ..., T-1$, let $C_n = h(\cdot, n)^{-1}(0, +\infty)$, then we have that if $a > b > \log\frac{\pi_0}{\pi_1}$ and $a \in C_n$, then $b \in C_n$; if $a < b < \log\frac{\pi_0}{\pi_1}$ and $a \in C_n$, then $b \in C_n$. Furthermore, $C_{n+1} \subset C_n \subset C$, where C is defined in (S1.8).

With the aid of Lemma 3, Theorem 1 can be proved similarly as that of Proposition 1. We omit the details.

S1.3 Proof of Theorem 2

For a truncated test with truncation length T, we consider the minimal conditional risk with n samples

$$V_n^T(x_{1:n}, j_{1:n}) = \inf_{(J,N,D)\in\mathcal{A}^T_{x_{1:n},j_{1:n}}} E\Big[L\{(N,D),\theta\}\Big|X_{1:n}=x_{1:n}\Big].$$

According to Lemma 3, $V_n^T(x_{1:n}, j_{1:n})$ depends only on the log-likelihood ratio statistic l that is defined in (S1.7). We abuse the notation a little and write

$$V_n^T(a) = \inf_{(J,N,D)\in\mathcal{A}^T_{x_{1:n},j_{1:n}}} E\Big[L\{(N,D),\theta\}\Big|l(X_{1:n}, j_{1:n})=a\Big].$$

Because $\mathcal{A}^T_{x_{1:n},j_{1:n}}$ is increasing in T, so $V_n^T(a)$ is non-increasing in T for all $n=0,1,2,...$ and $a\in\mathbb{R}$. We write $V_n^\infty(a)=\lim_{T\to\infty}V_n^T(a)$, for each $a\in\mathbb{R}$. For each T, $V_n^T(a)$ follows the Bellman equation

$$V_n^T(a) = \min\left\{\Phi_n(a), \inf_{\delta_{n+1}}\mathbb{E}\left[V_{n+1}^T\left(l+\log\frac{f_{1,\delta_{n+1}}(X_{n+1})}{f_{0,\delta_{n+1}}(X_{n+1})}\right)\ \Big|\ l(X_{1:n}, j_{1:n})=a\right]\right\}, \quad \text{(S1.15)}$$

where $\Phi_n(a)$ is the minimal conditional risk for stopping with n samples

$$\Phi_n(a) = \min\{\frac{\pi_0}{\pi_0+\pi_1 e^a}, \frac{\pi_1 e^a}{\pi_0+\pi_1 e^a}\} + nc.$$

Let $T\to\infty$ in (S1.15) and by monotone convergence theorem, we have

$$V_n^\infty(a) = \min\left\{\Phi_n(a), \inf_{\delta_n+1}\mathbb{E}\left[V_{n+1}^\infty\left(a+\log\frac{f_{1,\delta_{n+1}}(X_{n+1})}{f_{0,\delta_{n+1}}(X_{n+1})}\right)|l(X_{1:n}, j_{1:n})=a\right]\right\}. \quad \text{(S1.16)}$$

Let (J^*, N^*, D^*) be the optimal non-truncated test procedure that is defined in (4.16). According to Proposition 1, there exists experiment selection function j^* such that $j^*_{n+1}(X_{1:n}) = j^*(l(X_{1:n}, j^*_{1:n}))$. Let $\delta^*_{n+1} = j^*(l(X_{1:n}, j^*_{1:n}))$ be the stochastic process of experiment selection.

We define the following stochastic process

$$W_n = V_n^{\infty}(l(X_{1:n}, j_{1:n}^*)).$$

According to (S1.16), the process $\{W_n : n \geq 0\}$ is a sub-martingale with respect to the filtration $\mathcal{G}_n = \sigma(l_m^*, m \leq n)$, where we define the stochastic process $l_m^* = l(X_{1:m}, j_{1:m}^*)$. To see why $\{W_n : n \geq 0\}$ is a sub-martingale,

$$\begin{aligned} W_n = V_n^{\infty}(l_n^*) \leq & \inf_{\delta_{n+1}} \mathbb{E}\left[V_{n+1}^{\infty}\left(l_n^* + \log\frac{f_{1,\delta_{n+1}}(X_{n+1})}{f_{0,\delta_{n+1}}(X_{n+1})}\right) \Bigm| l_n^*\right] \\ \leq & \mathbb{E}\left[V_{n+1}^{\infty}\left(l_n^* + \log\frac{f_{1,j_{n+1}^*(X_{1:n})}(X_{n+1})}{f_{0,j_{n+1}^*(X_{1:n})}(X_{n+1})}\right) \Bigm| l_n^*\right] \\ = & \mathbb{E}\left[V_{n+1}^{\infty}(l_{n+1}^*)|l_n^*\right] = \mathbb{E}(W_{n+1}|\mathcal{G}_n). \end{aligned}$$

Note that $\{W_{n\wedge N^*} : n = 1, 2, ...\}$ is uniformly integrable, where $n \wedge N^* = \min(n, N^*)$. Using optional stopping theorem, we have

$$\mathbb{E}[W_{N^*}] \geq W_0 = V_0^{\infty}(0). \tag{S1.17}$$

According to (S1.16), we have $W_{N^*} \leq \Phi_{N^*}(l_{N^*}^*)$. The above display together with (S1.17) gives

$$\mathbb{E}[\Phi_{N^*}(l_{N^*}^*)] \geq V_0^{\infty}(0).$$

Note that $\mathbb{E}[\Phi_{N^*}(l_{N^*}^*)] = \min_{(J,N,D)\in\mathcal{A}} \mathbf{R}(J, N, D)$ and $V_0^{\infty}(0) = \lim_{T\to\infty} \min_{(J,N,D)\in\mathcal{A}^T} \mathbf{R}(J, N, D)$. Consequently,

$$\lim_{T\to\infty} \min_{(J,N,D)\in\mathcal{A}^T} \mathbf{R}(J, N, D) \leq \min_{(J,N,D)\in\mathcal{A}} \mathbf{R}(J, N, D). \tag{S1.18}$$

The converse inequality is obvious. Since for any T, $\mathcal{A}^T \subseteq \mathcal{A}$,

$$\min_{(J,N,D)\in\mathcal{A}^T} \mathbf{R}(J, N, D) \geq \min_{(J,N,D)\in\mathcal{A}} \mathbf{R}(J, N, D),$$

which implies that,

$$\lim_{T\to\infty}\min_{(J,N,D)\in\mathcal{A}^T}\mathbf{R}(J,N,D)\geq\min_{(J,N,D)\in\mathcal{A}}\mathbf{R}(J,N,D). \tag{S1.19}$$

We complete the proof by combining (S1.18) and (S1.19).

S1.4 Proof of Theorem 3

We first define the filtration $\mathcal{F}_k$ as the σ-field generated by both the $\theta_1,...,\theta_k$ and the observations $X_{1,1:N_1},...,X_{k,1:N_k}$, where $X_{k,1:N_k}$ denotes the responses to object k. In addition, let

$$Y_k=\mathbb{E}\Big[L((N_k,D_k),\theta_k)|\mathcal{F}_{k-1}\Big],$$

where the loss function L is defined in (S1.1). Note that θ_k is independent with $\mathcal{F}_{k-1}$. Therefore,

$$Y_k=\widetilde{R}(\pi_1,\widehat{\pi}_1^{(k-1)}),$$

where

$$\begin{aligned}\widetilde{R}(\pi_1,\widehat{\pi}_1^{(k)}) &= \pi_0\mathbb{P}(D_k=1|\widehat{\pi}_1^{(k-1)},\theta_k=0)+\pi_1\mathbb{P}(D_k=0|\widehat{\pi}_1^{(k-1)},\theta_k=1)\\ &\quad +c\pi_0\mathbb{E}(N_k|\widehat{\pi}_1^{(k-1)},\theta_k=0)+c\pi_1\mathbb{E}(N_k|\widehat{\pi}_1^{(k-1)},\theta_k=1). \end{aligned}\tag{S1.20}$$

We notice the that $c\leq\widehat{\pi}_1^{(k-1)}\leq 1-c$, so the conditional expectations $\mathbb{E}(N_k|\widehat{\pi}_1^{(k-1)},\theta_k=0)$ and $\mathbb{E}(N_k|\widehat{\pi}_1^{(k-1)},\theta_k=1)$ are bounded. Also notice that $\widetilde{R}$ is a linear function in π_1 and thus Lipschitz in π_1, so there exists a positive number κ_1 such that

$$|\widetilde{R}(\pi_1,\widehat{\pi}_1^{(k-1)})-\widetilde{R}(\widehat{\pi}_1^{(k-1)},\widehat{\pi}_1^{(k-1)})|\leq\kappa_1|\pi_1-\widehat{\pi}_1^{(k-1)}|. \tag{S1.21}$$

Because $\widehat{\pi}^{(k-1)}$ is consistent and (S1.21), we have

$$\widetilde{R}(\pi_1, \widehat{\pi}_1^{(k-1)}) - \widetilde{R}(\widehat{\pi}_1^{(k-1)}, \widehat{\pi}_1^{(k-1)}) \xrightarrow{k\to\infty} 0 \quad \text{in probability.}$$

The next lemma shows that $\min \mathbf{R}(J, N, D)$ is also continuous in π_1. The proof for Lemma 4 is given in Section S2.

Lemma 4. *Let $\bar{R}(\pi_1) = \min \mathbf{R}(J, N, D)$ be the minimal Bayes risk corresponding to the prior probability $(1 - \pi_1, \pi_1)$, then the function $\bar{R}(\pi_1)$ is continuous with respect to π_1. In addition, there exists a positive constant κ_2 such that for all $c \le \pi_1, \pi_1' \le 1 - c$*

$$|\bar{R}(\pi_1) - \bar{R}(\pi_1')| \le \kappa_2 |\pi_1 - \pi_1'| \tag{S1.22}$$

Note that $\widetilde{R}(\widehat{\pi}_1^{(k-1)}, \widehat{\pi}_1^{(k-1)}) = \bar{R}(\widehat{\pi}_1^{(k-1)})$ and $\bar{R}(\pi_1) = \min \mathbf{R}(J, N, D)$. By the continuity of $\bar{R}(\pi_1)$ in Lemma 4 and the assumption $\widehat{\pi}^{(k-1)} \to \pi_1$ in probability, we have

$$\widetilde{R}(\pi_1, \widehat{\pi}_1^{(k-1)}) - \min \mathbf{R}(J, N, D) \xrightarrow{k\to\infty} 0 \quad \text{in probability.}$$

Furthermore, according to (S1.21) and (S1.22),

$$|\widetilde{R}(\pi_1, \widehat{\pi}_1^{(k-1)}) - \min \mathbf{R}(J, N, D)| \le (\kappa_1 + \kappa_2) |\widehat{\pi}_1^{(k-1)} - \pi_1| \le \kappa_1 + \kappa_2.$$

The above display together with the dominated convergence theorem imply that

$$\lim_{k\to\infty} \mathbb{E} |\widetilde{R}(\pi_1, \widehat{\pi}_1^{(k-1)}) - \min \mathbf{R}(J, N, D)| = 0.$$

Consequently,

$$\lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \mathbb{E} |\widetilde{R}(\pi_1, \widehat{\pi}_1^{(k-1)}) - \min \mathbf{R}(J, N, D)| = 0. \tag{S1.23}$$

For any $\varepsilon > 0$, we apply the Chebyshev's inequality and obtain

$$\mathbb{P}\Big(|\frac{1}{K}\sum_{k=1}^{k}\widetilde{R}(\pi_1, \widehat{\pi}_1^{(k-1)}) - \min \mathbf{R}(J,N,D)| > \varepsilon\Big) \le \frac{1}{\varepsilon K}\sum_{k=1}^{K}\mathbb{E}|\widetilde{R}(\pi_1, \widehat{\pi}_1^{(k-1)}) - \min \mathbf{R}(J,N,D)|.$$

Recall $Y_k = \widetilde{R}(\pi_1, \widehat{\pi}_1^{(k-1)})$, then the above inequality and (S1.23) give

$$\frac{1}{K}\sum_{k=1}^{K} Y_k - \min \mathbf{R}(J,N,D) \xrightarrow{K\to\infty} 0 \quad \text{in probability.} \tag{S1.24}$$

We proceed to the limit of $L_K = \frac{1}{K}\sum_{k=1}^{K} L\{(N_k, D_k), \theta_k\}$. Note that

$$\mathbb{E}\Big[L\{(N_k, D_k), \theta_k\}|\mathcal{F}_{k-1}\Big] = Y_k.$$

Consequently, $\sum_{k=1}^{K} L\{(N_k, D_k), \theta_k\} - Y_k$ is a martingale with respect to the filtration $\{\mathcal{F}_K : K = 1, 2, ...\}$. Standard calculation for square integrable martingale yields

$$\mathbb{E}\Big[\sum_{k=1}^{K} L\{(N_k, D_k), \theta_k\} - Y_k\Big]^2 = \sum_{k=1}^{K}\mathbb{E}[L\{(N_k, D_k), \theta_k\} - Y_k]^2 \le \kappa_3 K.$$

for some positive constant κ_3. We apply Chebyshev's inequality to the above display

$$\mathbb{P}\Big(|L_K - \frac{1}{K}\sum_{k=1}^{K} Y_k| > \varepsilon\Big) \le \frac{1}{K^2\varepsilon^2}\mathbb{E}\Big[\sum_{k=1}^{K} L\{(N_k, D_k), \theta_k\} - Y_k\Big]^2 \le \frac{\kappa_3}{K\varepsilon^2}$$

for an arbitrary positive constant ε. This implies that

$$L_K - \frac{1}{K}\sum_{k=1}^{K} Y_k \xrightarrow{K\to\infty} 0 \qquad \text{in probability.} \tag{S1.25}$$

We complete the proof by combining (S1.25) and (S1.24).

S2 Proof of Supporting Lemmas

S2.1 Proof of Lemma 1

It is sufficient to show that if

$$l(x_{1:n}, j_{1:n}) = l(\bar{x}_{1:\bar{n}}, \bar{j}_{1:\bar{n}}), \tag{S2.26}$$

then $g(x_{1:n}, j_{1:n}) = g(\bar{x}_{1:\bar{n}}, \bar{j}_{1:\bar{n}})$. If in the contrary, assume without loss of generality that $g(x_{1:n}, j_{1:n}) > g(\bar{x}_{1:\bar{n}}, \bar{j}_{1:\bar{n}})$, then according to the definition of g, there exist $(J, N, D) \in \mathcal{A}_{x_{1:n}, j_{1:n}}$ such that

$$r_s(x_{1:n}, j_{1:n}) - \mathbb{E}^J \left[L\{(N, D), \theta\} \middle| X_{1:n} = x_{1:n} \right] > g(\bar{x}_{1:\bar{n}}, \bar{j}_{1:\bar{n}}).$$

We use the superscript J in the expectation sign to indicate the expectation is computed with the experiment selection rule J. We construct a sequential adaptive design $(\bar{J}, \bar{N}, \bar{D}) \in \mathcal{A}_{\bar{x}_{1:\bar{n}}, \bar{j}_{1:\bar{n}}}$ as follows. For any observations

$$\bar{x}_1, \bar{x}_2, ..., \bar{x}_{\bar{n}}, y_1, y_2, ...$$

we first choose the experiment selection function

$$\bar{j}_{\bar{n}+m+1}(\bar{x}_{1:\bar{n}}, y_{1:m}) = j_{n+m+1}(x_{1:n}, y_{1:m}).$$

Next, for $m = 1, 2, ...,$ to decide whether the test procedure $(\bar{J}, \bar{N}, \bar{D})$ stops or not with observations

$$\bar{x}_1, ..., \bar{x}_{\bar{n}}, y_1, ..., y_m,$$

we look at if (J, N, D) stop with observations

$$x_1, ..., x_n, y_1, ..., y_m$$

or not. If (J, N, D) stops with observations $x_{1:n}, y_{1:m}$ then we let $(\bar{J}, \bar{N}, \bar{D})$ stop with observations $\bar{x}_{1:\bar{n}}, y_{1:m}$, and otherwise we let the test $(\bar{J}, \bar{N}, \bar{D})$ do not stop. Lastly, for the decision $\bar{D}$ with observations $\bar{x}_{1:\bar{n}}, y_{1:m}$, we also let it make the same decision as that of D with observations $x_{1:n}, y_{1:m}$. In short, we let the sequential adaptive design $(\bar{J}, \bar{N}, \bar{D})$ do whatever the test procedure (J, N, D) do by replacing the first $\bar{n}$ observations with $x_{1:n}$.

We consider the reduced conditional risk for $(\bar{J}, \bar{N}, \bar{D})$,

$$r_s(x_{1:\bar{n}}, j_{1:\bar{n}}) - \mathbb{E}^{\bar{J}}\Big[L\{(\bar{N}, \bar{D}), \theta\}\Big| X_{1:\bar{n}} = \bar{x}_{1:\bar{n}}\Big]. \tag{S2.27}$$

Notice that for any possible sequence of observations

$$\bar{x}_1, ..., \bar{x}_{\bar{n}}, y_1, y_2, ...$$

and

$$x_1, ..., x_n, y_1, y_2, ...$$

The decision $\bar{D} = D$, and the stopping time

$$\bar{N} - \bar{n} = N - n.$$

In addition, the posterior distribution of $X_{n+1}, X_{n+2}, ...$ and $X_{\bar{n}+1}, X_{\bar{n}+2}, ...$ are the same with the same experiment selection rule J and $\bar{J}$ for future experiments conditional on $X_{1:n} = x_{1:n}$ and $X_{1:\bar{n}} = \bar{x}_{1:\bar{n}}$ respectively. To see this point, notice that the conditional distribution $X_{n+1}|\theta, X_{1:\bar{n}} = \bar{x}_{1:\bar{n}}$ has the density function $f_{\theta, \bar{j}_{n+1}(\bar{x}_{1:\bar{n}})}(X_{n+1})$ with the experiment selection rule $\bar{J}$. Since $\bar{j}_{n+1}(\bar{x}_{1:\bar{n}}) = j_{n+1}(x_{1:n})$ by our construction, $f_{\theta,\bar{j}_{n+1}(\bar{x}_{1:\bar{n}})}(X_{n+1}) = f_{\theta,j_{n+1}(x_{1:n})}(X_{n+1})$, which implies that $X_{n+1}|\theta, X_{1:\bar{n}} = \bar{x}_{1:\bar{n}}$ has the same conditional distribution using the experiment selection rule J as $X_{n+1}|\theta, X_{1:n} = x_{1:n}$. The above claim directly follows by an induction argument. Therefore, by (S2.26), for any given m, we have the same conditional distribution for the sequence $X_{n+1:n+m}|\theta, X_{1:n} = x_{1:n}$ with selection rule $\bar{J}$ and $X_{\bar{n}+1:\bar{n}+m}|\theta, X_{1:\bar{n}} = \bar{x}_{1:\bar{n}}$ with J. Furthermore, the posterior distributions of θ are the same given $X_{1:n} = x_{1:n}$ and $X_{1:\bar{n}} = \bar{x}_{1:\bar{n}}$ with selection rule J and $\bar{J}$ respectively. Thus, we have

$$\mathbb{E}^{\bar{J}}\Big[L\{(\bar{N},\bar{D}),\theta\}\Big|X_{1:\bar{n}} = \bar{x}_{1:\bar{n}}\Big] - \bar{n}c = \mathbb{E}^{J}\Big[L\{(N,D),\theta\}\Big|X_{1:n} = x_{1:n}\Big] - nc. \tag{S2.28}$$

Recall that

$$r_s(x_{1:\bar{n}}, j_{1:\bar{n}}) = \min\{\frac{\pi_0}{\pi_0 + \pi_1 e^{l(\bar{x}_{1:\bar{n}}, j_{1:\bar{n}})}}, \frac{\pi_1 e^{l(\bar{x}_{1:\bar{n}}, j_{1:\bar{n}})}}{\pi_0 + \pi_1 e^{l(\bar{x}_{1:\bar{n}}, j_{1:\bar{n}})}}\} + \bar{n}c,$$
$$r_s(x_{1:n}, j_{1:n}) = \min\{\frac{\pi_0}{\pi_0 + \pi_1 e^{l(x_{1:n}, j_{1:n})}}, \frac{\pi_1 e^{l(x_{1:n}, j_{1:n})}}{\pi_0 + \pi_1 e^{l(x_{1:n}, j_{1:n})}}\} + nc.$$

Further, by (S2.26),

$$r_s(\bar{x}_{1:\bar{n}}, j_{1:\bar{n}}) - \bar{n}c = r_s(x_{1:n}, j_{1:n}) - nc$$

The above display together with (S2.28) implies

$$\begin{aligned}
&g(\bar{x}_{1:\bar{n}}, \bar{j}_{1:\bar{n}})\\
&\geq r_s(x_{1:\bar{n}}, j_{1:\bar{n}}) - \mathbb{E}^{\bar{J}}\Big[L\{(\bar{N},\bar{D}),\theta\}\Big|X_{1:\bar{n}} = \bar{x}_{1:\bar{n}}\Big]\\
&= r_s(x_{1:n}, j_{1:n}) - \mathbb{E}^{J}\Big[L\{(N,D),\theta\}\Big|X_{1:n} = x_{1:n}\Big]\\
&> g(\bar{x}_{1:\bar{n}}, \bar{j}_{1:\bar{n}})
\end{aligned}$$

which contradicts with the assumption that $g(x_{1:n}, j_{1:n}) > g(\bar{x}_{1:\bar{n}}, \bar{j}_{1:\bar{n}})$.

S2.2 Proof of Lemma 2

For $a > b > \log \frac{\pi_0}{\pi_1}$, let $(x_{1:n}, j_{1:n})$ and $(\bar{x}_{1:\bar{n}}, \bar{j}_{1:\bar{n}})$ be such that $l(x_{1:n}, j_{1:n}) = a$ and $l(\bar{x}_{1:\bar{n}}, \bar{j}_{1:\bar{n}}) = b$. We assume that $g(x_{1:n}, j_{1:n}) > 0$. For the rest of the proof, we are going to show

$$g(\bar{x}_{1:\bar{n}}, \bar{j}_{1:\bar{n}}) > 0.$$

We use the similar method as in the proof of Lemma 1. $g(x_{1:n}, j_{1:n}) > 0$ implies that there exists $(J, N, D) \in \mathcal{A}_{x_{1:n}, j_{1:n}}$ such that

$$r_s(x_{1:n}, j_{1:n}) - \mathbb{E}^J\left[L\{(N, D), \theta\}\middle|X_{1:n} = x_{1:n}\right] > 0 \tag{S2.29}$$

Now we construct the sequential adaptive design $(\bar{J}, \bar{N}, \bar{D}) \in \mathcal{A}_{\bar{x}_{1:\bar{n}}, \bar{j}_{1:\bar{n}}}$ the same way as that in the proof of Lemma 1. Using the same arguments as in the proof of Lemma 1, we have

$$\begin{aligned} E_0 :=& \mathbb{E}^J\left[L\{(N, D), \theta\}\middle|X_{1:n} = x_{1:n}, \theta = 0\right] - nc \\ =& \mathbb{E}^{\bar{J}}\left[L\{(\bar{N}, \bar{D}), \theta\}\middle|X_{1:\bar{n}} = \bar{x}_{1:\bar{n}}, \theta = 0\right] - \bar{n}c, \end{aligned} \tag{S2.30}$$

and

$$\begin{aligned} E_1 :=& \mathbb{E}^J\left[L\{(N, D), \theta\}\middle|X_{1:n} = x_{1:n}, \theta = 1\right] - nc \\ =& \mathbb{E}^{\bar{J}}\left[L\{(\bar{N}, \bar{D}), \theta\}\middle|X_{1:\bar{n}} = \bar{x}_{1:\bar{n}}, \theta = 1\right] - \bar{n}c. \end{aligned} \tag{S2.31}$$

Notice that $b > \log \frac{\pi_0}{\pi_1}$ and $l(\bar{x}_{1:\bar{n}}, \bar{j}_{1:\bar{n}}) = b$. Consequently,

$$r_s(\bar{x}_{1:\bar{n}}, \bar{j}_{1:\bar{n}}) = \frac{\pi_0}{\pi_0 + \pi_1 e^b} + \bar{n}c. \tag{S2.32}$$

We combine (S2.30), (S2.31) and (S2.32), and arrive at

$$\begin{aligned} & r_s(x_{1:\bar{n}}, j_{1:\bar{n}}) - \mathbb{E}^{\bar{J}}\left[L\{(\bar{N},\bar{D}),\theta\}\middle| X_{1:\bar{n}} = \bar{x}_{1:\bar{n}}\right] \\ =& \frac{\pi_0}{\pi_0 + \pi_1 e^b} - \mathbb{P}(\theta = 0 | X_{1:\bar{n}} = \bar{x}_{1:\bar{n}}) \times E_0 - \mathbb{P}(\theta = 1 | X_{1:\bar{n}} = \bar{x}_{1:\bar{n}}) \times E_1 \\ =& \frac{\pi_0}{\pi_0 + \pi_1 e^b} - \frac{\pi_0}{\pi_0 + \pi_1 e^b} \times E_0 - \frac{\pi_1 e^b}{\pi_0 + \pi_1 e^b} \times E_1 \\ =& \frac{\pi_0(1 - E_0) - \pi_1 e^b E_1}{\pi_0 + \pi_1 e^b}. \end{aligned} \tag{S2.33}$$

Similarly, we have

$$r_s(x_{1:n}, j_{1:n}) - \mathbb{E}^{J}\left[L\{(N,D),\theta\}\middle| X_{1:n} = x_{1:n}\right] = \frac{\pi_0(1 - E_0) - \pi_1 e^a E_1}{\pi_0 + \pi_1 e^a}.$$

According to (S2.29) and the above display, we have

$$\frac{\pi_0(1 - E_0) - \pi_1 e^a E_1}{\pi_0 + \pi_1 e^a} > 0, \tag{S2.34}$$

which implies that

$$\pi_0(1 - E_0) - \pi_1 e^a E_1 > 0.$$

Because $\pi_0(1 - E_0) - \pi_1 e^b E_1 > \pi_0(1 - E_0) - \pi_1 e^a E_1$ and (S2.34), we have

$$\frac{\pi_0(1 - E_0) - \pi_1 e^b E_1}{\pi_0 + \pi_1 e^b} > 0.$$

According to the above display, the definition of g and (S2.33), we have

$$\begin{aligned} g(\bar{x}_{1:\bar{n}}, \bar{j}_{1:\bar{n}}) \geq & r_s(x_{1:\bar{n}}, j_{1:\bar{n}}) - \mathbb{E}^{\bar{J}}\left[L\{(\bar{N},\bar{D}),\theta\}\middle| X_{1:\bar{n}} = \bar{x}_{1:\bar{n}}\right] \\ \geq & \frac{\pi_0(1 - E_0) - \pi_1 e^b E_1}{\pi_0 + \pi_1 e^b} > 0. \end{aligned}$$

With similar arguments, if $a < b < \log \frac{\pi_0}{\pi_1}$ and $h(a) > 0$, then we have $h(b) > 0$. We omit the details.

S2.3 Proof of Lemma 3

The proof of the first half of the Lemma is similar to that of Lemma 2, and is thus omitted. That is, there exists $h^\dagger$ satisfying (S1.14), and for each C_n, if $a > b > \log \frac{\pi_0}{\pi_1}$ and $a \in C_n$ then, $b \in C_n$. We proceed to prove that

$$C_n \subset C_{n-1}.$$

It is sufficient to show that for each $a \in C_{n+1}$, we also have $a \in C_n$. Due to the symmetry of the problem, we focus on the case where $a > \log \frac{\pi_0}{\pi_1}$. Let $\bar{n} = n - 1$ and let $(x_{1:n}, j_{1:n})$ and $(\bar{x}_{1:\bar{n}}, \bar{j}_{1:\bar{n}})$ be such that $l(x_{1:n}, j_{1:n}) = a$ and $l(\bar{x}_{1:\bar{n}}, \bar{j}_{1:\bar{n}}) = a$. We assume that $g^\dagger(x_{1:n}, j_{1:n}) > 0$. For the rest of the proof, we are going to show

$$g^\dagger(\bar{x}_{1:\bar{n}}, \bar{j}_{1:\bar{n}}) > 0.$$

We use the similar method as in the proof of Lemma 1. Note that $g^\dagger(x_{1:n}, j_{1:n}) > 0$ implies that there exists $(J, N, D) \in \mathcal{A}^T_{x_{1:n}, j_{1:n}}$ such that

$$r_s(x_{1:n}, j_{1:n}) - \mathbb{E}^J \left[L\{(N, D), \theta\} \middle| X_{1:n} = x_{1:n} \right] > 0, \tag{S2.35}$$

where $\mathcal{A}^T_{x_{1:n}, j_{1:n}}$ is defined similar to $\mathcal{A}_{x_{1:n}, j_{1:n}}$ but requires that $N \leq T$. Now we construct the sequential adaptive design $(\bar{J}, \bar{N}, \bar{D})$ the same way as that in the proof of Lemma 1.

Because $\bar{n} = n + 1 > n$, from the construction, we have $\bar{N} = \bar{N} - \bar{n} + \bar{n} = N - n + \bar{n} = N - n + n - 1 = N - 1 \leq T$. Thus, $(\bar{J}, \bar{N}, \bar{D}) \in \mathcal{A}_{\bar{x}^T_{1:\bar{n}}, \bar{j}_{1:\bar{n}}}$. Using the same arguments as in the proof of Lemma 2, we can see that

$$r_s(\bar{x}_{1:\bar{n}}, \bar{j}_{1:\bar{n}}) = \frac{\pi_0}{\pi_0 + \pi_1 e^a} + \bar{n}c. \tag{S2.36}$$

On the other hand, from the construction,

$$\mathbb{E}^{J}\left[L\{(N,D),\theta\}\Big|X_{1:n}=x_{1:n}\right]-nc=\mathbb{E}^{\bar{J}}\left[L\{(\bar{N},\bar{D}),\theta\}\Big|X_{1:\bar{n}}=\bar{x}_{1:\bar{n}}\right]-\bar{n}c. \tag{S2.37}$$

Combining (S2.35), (S2.36) and (S2.37), we can see that $g^{\dagger}(\bar{x}_{1:\bar{n}},\bar{j}_{1:\bar{n}})>0$. Therefore, $a\in C_{\bar{n}}=C_{n-1}$. This completes our proof.

S2.4 Proof of Lemma 4

We consider the Bayes risk when the prior probability is $(1-\pi_1,\pi_1)$,

$$\begin{aligned}\mathbf{R}^{\pi_1}(J,N,D)=&(1-\pi_1)\mathbb{P}(D=1|\theta=0)+\pi_1\mathbb{P}(D=0|\theta=1)\\&+c\{\pi_0\mathbb{E}(N|\theta=0)+\pi_1\mathbb{E}(N|\theta=1)\}.\end{aligned}$$

Here we use the superscript π_1 to indicate the prior. For fixed (J,N,D) the function $\mathbf{R}^{\pi_1}(J,N,D)$ is linear in π_1, and is thus continuous in π_1. Let $(J^{\pi_1},N^{\pi_1},D^{\pi_1})$ be the optimal procedure for the prior probability $\mathbb{P}(\theta=1)=\pi_1$. Then,

$$\mathbf{R}^{\pi_1}(J^{\pi_1},N^{\pi_1},D^{\pi_1})=\min\mathbf{R}^{\pi_1}(J,N,D)=\bar{R}(\pi_1).$$

Now we consider two prior probability π_1 and $\widetilde{\pi}_1$. We have

$$\begin{aligned}\bar{R}(\pi_1)-\bar{R}(\widetilde{\pi}_1)&=\min\mathbf{R}^{\pi_1}(J,N,D)-\mathbf{R}^{\widetilde{\pi}_1}(J^{\widetilde{\pi}_1},N^{\widetilde{\pi}_1},D^{\widetilde{\pi}_1})\\&\leq\mathbf{R}^{\pi_1}(J^{\widetilde{\pi}_1},N^{\widetilde{\pi}_1},D^{\widetilde{\pi}_1})-\mathbf{R}^{\widetilde{\pi}_1}(J^{\widetilde{\pi}_1},N^{\widetilde{\pi}_1},D^{\widetilde{\pi}_1})\end{aligned}$$

and similarly,

$$\bar{R}(\widetilde{\pi}_1)-\bar{R}(\pi_1)\leq\mathbf{R}^{\widetilde{\pi}_1}(J^{\pi_1},N^{\pi_1},D^{\pi_1})-\mathbf{R}^{\pi_1}(J^{\pi_1},N^{\pi_1},D^{\pi_1}).$$

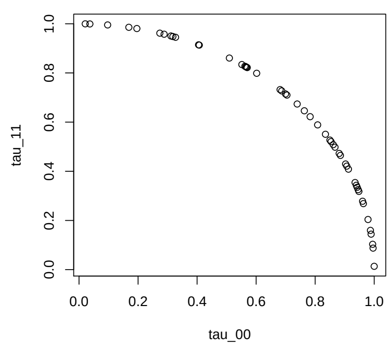


Figure 1: The quality parameters for 50 simulated workers

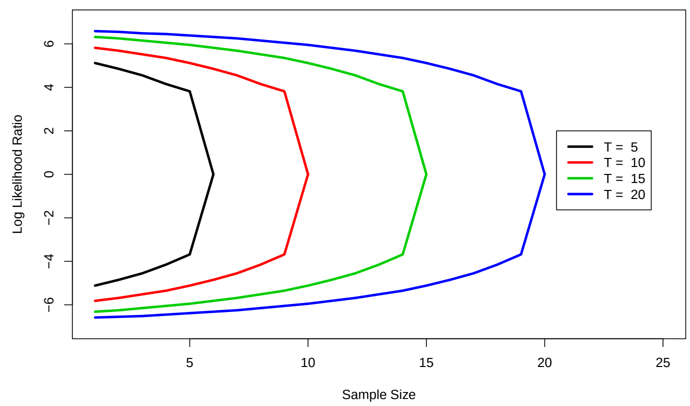


Figure 2: Hitting boundaries for different truncation lengths.

Furthermore, for all $\pi \in [c, 1-c]$ the conditional expectations $E(N^{\pi_1}|\theta = 0)$ and $E(N^{\pi_1}|\theta = 0)$ are bounded by some positive number κ_2. Therefore, the continuity of $\mathbf{R}^{\pi_1}(J, N, D)$ in π_1 implies the continuity of $\bar{R}(\pi_1)$, and we have

$$|\bar{R}(\widetilde{\pi}_1) - \bar{R}(\pi_1)| \leq \kappa_2 |\widetilde{\pi}_1 - \pi_1|.$$

Table 1: Performance of Ada-SPRT for different truncation lengths.

	$T = 5$	$T = 10$	$T = 15$	$T = 20$
Stopping Time	5.000	9.276	12.301	14.505
Accuracy	0.857	0.926	0.961	0.977
Loss	14.468	7.672	4.240	2.630

S3 Simulated Experiments

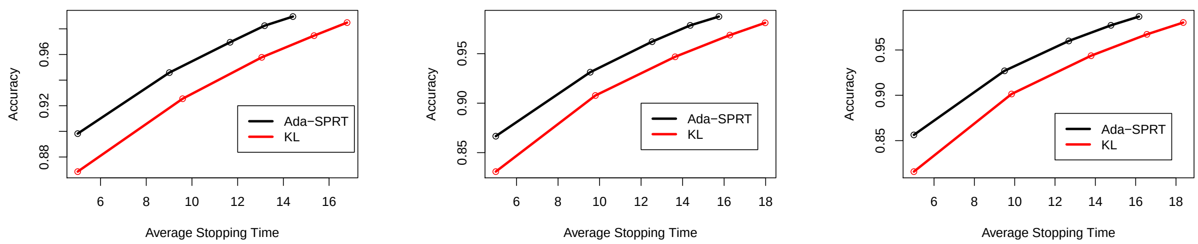
S3.1 Effect of Truncation Length T

We first study the effect of the truncation length T for a single hypothesis. We simulate $M = 50$ workers with quality parameters for worker i:

$$\gamma^i \sim \text{Uniform}(0, \frac{\pi}{2}),$$
$$\tau_{00}^i = \sin(\gamma^i), \quad \tau_{11}^i = \cos(\gamma^i).$$

A scatter plot of the generated τ_{00}^i for $1 \le i \le M$ is shown in Figure 1. We generate 50 workers in this way such that no worker is dominantly worse than another. That is, there does not exist a pair of workers i and i' such that $\tau_{00}^i < \tau_{00}^{i'}$ and $\tau_{11}^i < \tau_{11}^{i'}$.

We consider a single hypothesis testing problem (i.e., labeling for a single object) with the true label θ drawn from the Bernoulli distribution with $\pi_0 = \pi_1 = 0.5$. In this experiment, since our main goal is to investigate the effect of truncation length T, we assume true π_1 and workers' parameters are known for simplicity and set the parameter $c = 2^{-12}$. We vary the truncation length $T = 5, 10, 15$, and 20. For different truncation lengths, we plot the hitting



(a) $\pi_1 = 0.8$ (b) $\pi_1 = 0.65$ (c) $\pi_1 = 0.5$

Figure 3: Comparison between the Ada-SPRT and KL approaches.

boundaries in Figure 2. As one can see, given any fixed truncation length T, for different sample sizes from 1 to T (on the x-axis of Figure 2), we have

$$B^\dagger(1) \leq B^\dagger(2) \leq ... \leq B^\dagger(T) = \log \frac{\pi_0}{\pi_1} = 0 = A^\dagger(T) \leq A^\dagger(T-1) \leq ... \leq A^\dagger(1).$$

This observation is consistent with our result in Theorem 1.

Now for each truncation length T, we generate 50,000 independent replications and run Ada-SPRT for each replication. In Table 1, we report the *average* of (1) the stopping time N, (2) the labeling accuracy $\mathbf{1}_{\{D=\theta\}}$, and (3) the loss $\mathbf{1}_{\{D\neq\theta\}} + cN$ over 50,000 replications. As can be seen from Table 1, as the truncation length increases, both the stopping time and accuracy increase simultaneously. However, the average loss, which consists of labeling error and cost, decreases as T becomes larger.

S3.2 Comparison with the asymptotically optimal KL-information Approach

We compare the proposed Ada-SPRT procedure with an asymptotically optimal Kullback-Leibler (KL) approach from Chernoff (1959). The worker selection rule of the KL approach is based on workers' KL information, where the KL information for worker $\delta \in I$ given $\theta = 0$ and $\theta = 1$ is defined as

$$KL(0, \delta) = \mathbb{E}\left[\log \frac{f_{0,\delta}(X)}{f_{1,\delta}(X)} | \theta = 0\right], \quad \text{and} \quad KL(1, \delta) = \mathbb{E}\left[\log \frac{f_{1,\delta}(X)}{f_{0,\delta}(X)} | \theta = 1\right].$$

At time n, let $\pi(\theta = 0|l)$ and $\pi(\theta = 1|l)$ be the posterior probabilities under the current log-likelihood ratio l. Then the worker selection rule of the KL approach is

$$j(l, n) = \begin{cases} \arg\max_{\delta \in I} KL(0, \delta), & \text{if } \pi(\theta = 0|l) > \pi(\theta = 1|l), \\ \arg\max_{\delta \in I} KL(1, \delta), & \text{otherwise.} \end{cases}$$

That is, the worker with the largest KL information at the posterior mode of θ is selected. In terms of the stopping rule, this KL approach adopts flat boundaries

$$A = -\log c + \log\left(\frac{\pi_0 \max_{\delta \in I} KL(1, \delta)}{\pi_1}\right) \quad \text{and} \quad B = \log c + \log\left(\frac{\pi_0}{\pi_1 \max_{\delta \in I} KL(0, \delta)}\right),$$

where the second terms in both A and B take the prior information and the worker pool quality into account. The algorithm stops once the log-likelihood ratio l crosses the boundaries, i.e., $l \geq A$ or $l \leq B$, or the sample size n has reached the truncation length T. The decision is based on the posterior probabilities upon stopping, that is, $D = \arg\max_{d \in \{0,1\}} \pi(\theta = d|l)$.

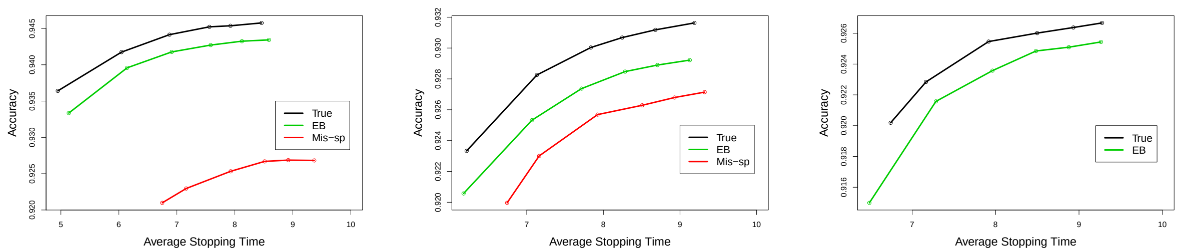
To compare the Ada-SPRT and KL approaches, the same worker pool in Section S3.1 is used. We consider three possible values of the class prior π_1: (1) $\pi_1 = 0.8$ (highly unbalanced class) (2) $\pi_1 = 0.65$ (moderately unbalanced class) (3) $\pi_1 = 0.5$ (balanced class). We set $c = 2^{-12}$ and vary the truncation length $T = 5, 10, 15, 20, 25$. For each π_1, c, and T, $500,000$ independent replications are generated. Results are summarized in Figure 3, where for each choice of π_1, we report the average accuracy as a function of average stopping time under varying truncation length T. According to Figure 3, the proposed Ada-SPRT method performs substantially better than the KL procedure under a finite sample setting.

S3.3 Class Prior and Empirical Bayes Estimator

In this simulated experiment, we consider the multiple hypotheses testing problem in Section 5, i.e., labeling multiple objects. In particular, we generate $K = 100$ objects with true label θ_k from the Bernoulli distributions with true class prior π_1. We consider three possible values of π_1: (1) $\pi_1 = 0.8$ (highly unbalanced class) (2) $\pi_1 = 0.65$ (moderately unbalanced class) (3) $\pi_1 = 0.5$ (balanced class). For each π_1, we compare three following procedures:

1. Ada-SPRT with true class prior π_1;
2. Ada-SPRT with empirical Bayes estimation of the class prior π_1 in Algorithm 2;
3. Ada-SPRT with the mis-specified class prior 0.5. Note that in the third case when $\pi_1 = 0.5$, it is the same as the Ada-SPRT with the true class prior.

We vary the cost parameter $c = 2^{-\rho}$ with $\rho = 7, 8, ..., 12$, which leads to different stopping times. For each choice of π_1, we report in Figure 4 the average accuracy as a function of average stopping time (i.e., $\frac{1}{K}\sum_{k=1}^{K} N_k$ where N_k is the stopping time for the k-th object) for truncated test with $T = 10$ (right panels) over 5,000 independent replications. As



(a) $\pi_1 = 0.8$ (b) $\pi_1 = 0.65$ (c) $\pi_1 = 0.5$

Figure 4: Performance of empirical Bayes estimation for different class priors.

can be seen from Figure 4, the performance of Ada-SPRT with empirical Bayes estimation is close to Ada-SPRT with true prior especially when the stopping time goes large. In addition, the performance of Ada-SPRT with empirical Bayes estimation achieves much better performance than Ada-SPRT with a mis-specified class prior, which demonstrates the effectiveness of using empirical Bayes estimation.

References

Chernoff, H. (1959). Sequential design of experiments. *The Annals of Mathematical Statistics*, 30(3):755–770.