

---

# Technical Appendices for “AdaDetectGPT: Adaptive Detection of LLM-Generated Text with Statistical Guarantees”

---

Anonymous Author(s)

Affiliation

Address

email

## 1 A Details on the analytic example in Section 3

2 In this section, we provide rigorous discussion about the analytic example presented in Section 3.  
3 Noted that

$$\begin{aligned}\mathbb{E}_{\tilde{X}_t \sim q, X_{<t} \sim p} \{w(\log q(X_t))\} &= q(1)w(\log q(1)) + q(0)w(\log q(0)), \\ \mathbb{E}_{X_{<t+1} \sim p} \{w(\log q(X_t))\} &= p_t(1)w(\log q(1)) + p_t(0)w(\log q(0)).\end{aligned}$$

4 It follows that

$$\begin{aligned}\mathbb{E}_{\tilde{X}_t \sim q, X_{<t} \sim p} \{w(\log q(X_t))\} - \mathbb{E}_{X_{<t+1} \sim p} \{w(\log q(X_t))\} \\ = (q(1) - p_t(1)) [w(\log q(1)) - w(\log q(0))].\end{aligned}$$

5 If  $w$  is an identity function, i.e.,  $w(x) = x$ , then the statistics (5) becomes

$$\frac{1}{L} \log \left( \frac{q(1)}{q(0)} \right) \sum_{t=1}^L (q(1) - p_t(1)).$$

6 In this case, (5) converges to zero as  $q \rightarrow 1/2$  regardless the distribution of  $p_t$ . However, if we  
7 consider adaptive witness function, the statistics in (5) becomes

$$\frac{1}{L} [w(\log q(1)) - w(\log q(0))] \sum_{t=1}^L (q(1) - p_t(1)).$$

8 When  $q(1) \neq 1/2$  (without generality, we assume  $q(1) = 1 - q(0) > 1/2$ ), there always exists a  
9 witness function  $w(z) = \mathbb{I} \left\{ z > \frac{\log q(1) + \log q(0)}{2} \right\}$  such that (5) becomes

$$\begin{aligned}\frac{1}{L} [\mathbb{I}\{\log q(1) > \log q(0)\} - \mathbb{I}\{\log q(0) > \log q(1)\}] \sum_{t=1}^L (q(1) - p_t(1)) \\ = \frac{1}{L} \sum_{t=1}^L (q(1) - p_t(1)) = q(1) - \frac{1}{L} \sum_{t=1}^L p_t(1),\end{aligned}$$

10 which is independent of the log ratio.

## 11 B Implementation details

### 12 B.1 Data for estimating witness function

13 In this part, we illustrate how we fetch external human and machined-generate text datasets for  
14 training witness function in our experiments.

When testing the performance of AdaDetectGPT on one dataset (e.g., Xsum), we recruit the remaining datasets (e.g., squad and writing) for training the witness function. This ensures the data for testing would not be included for training. In the white-box setting, since  $q^{(s)}$  is known, then the machine-generated text in the training datasets are generated by  $q^{(s)}$ . In the black-box setting, since  $q^{(s)}$  is unknown, the machine-generated text in the training datasets are generated by the surrogate models. In the case where the sampling model and scoring model is different and both using surrogate models, then the generated texts comes from the surrogate scoring model.

## B.2 Details: the witness function estimation

In this section we elaborate the estimation of witness function. Recall that the population version of objective function for estimating  $w$  is

$$\frac{\sum_t [\mathbb{E}_{X_{<t} \sim p, \tilde{X}_t \sim q_t} w(\log q_t(\tilde{X}_t | X_{<t}))] - \sum_t [\mathbb{E}_{X_{<t} \sim p, \tilde{X}_t \sim p_t} w(\log q_t(\tilde{X}_t | X_{<t}))]}{\sqrt{\sum_t \mathbb{E}_{X_{<t} \sim p} \text{Var}_{\tilde{X}_t \sim q_t} (w(\log q_t(\tilde{X}_t | X_{<t})))}} ,$$

and we replace the expectations  $\mathbb{E}_{X_{<t} \sim p}$  in both the numerator and denominator with their empirical average over the passages of human. This leads to the following plug-in estimated objective function:

$$\begin{aligned} & \frac{\sum_i \sum_t \mathbb{E}[w(\log q_t(\tilde{X}_t | X_{<t}^{(i)}))] - \sum_i \sum_t [w(\log q_t(X_t^{(i)} | X_{<t}^{(i)})]}{\sqrt{\sum_i \sum_t \text{Var}_{\tilde{X}_t \sim q_t} (w(\log q_t(\tilde{X}_t | X_{<t}^{(i)})))}} \\ & \approx \frac{\sqrt{2} \sum_i \sum_t \mathbb{E}[w(\log q_t(\tilde{X}_t | X_{<t}^{(i)}))] - \sqrt{2} \sum_i \sum_t [w(\log q_t(X_t^{(i)} | X_{<t}^{(i)})]}{\sqrt{\sum_i \sum_t \text{Var}_{\tilde{X}_t \sim q_t} (w(\log q_t(\tilde{X}_t | X_{<t}^{(i)}))) + \sum_i \sum_t \text{Var}(w(\log q_t(X_t^{(i)} | X_{<t}^{(i)})))}} , \end{aligned}$$

where  $\tilde{X}_t$  is sampled from the machine and the approximation holds because of Assumption 4. Besides, taking the expectation over  $\tilde{X}_t$  is time consuming, then we approximate  $\mathbb{E}[w(\log q_t(\tilde{X}_t | X_{<t}^{(i)}))]$  and  $\text{Var}_{\tilde{X}_t \sim q_t} (w(\log q_t(\tilde{X}_t | X_{<t}^{(i)})))$  with

$$\mathbb{E}[w(\log q_t(\tilde{X}_t | \tilde{X}_{<t}))] \text{ and } \text{Var}_{\tilde{X}_t \sim q_t} (w(\log q_t(\tilde{X}_t | \tilde{X}_{<t})))$$

where  $\tilde{X}_{<t}$  is sampled from LLM. They can be easily estimated by the Monte Carol method as we can utilize the accessible machine to generate texts. Specifically, we generate a text  $\tilde{X}^{(i)} \sim q$ . And they are estimated by:

$$\mathbb{E}[w(\log q_t(\tilde{X}_t^{(i)} | \tilde{X}_{<t}^{(i)}))] \text{ and } \text{Var}_{\tilde{X}_t^{(i)} \sim q_t} (w(\log q_t(\tilde{X}_t^{(i)} | \tilde{X}_{<t}^{(i)}))).$$

After these approximation, we lead to the following two-sample  $t$ -test type objective that eliminates the constant terms:

$$\begin{aligned} & \frac{\sum_i \sum_t [w(\log q_t(\tilde{X}_t^{(i)} | \tilde{X}_{<t}^{(i)}))] - \sum_i \sum_t [w(\log q_t(X_t^{(i)} | X_{<t}^{(i)})]}{\sqrt{\sum_i \sum_t \text{Var}_{\tilde{X}_t \sim q_t} (w(\log q_t(\tilde{X}_t | \tilde{X}_{<t}^{(i)}))) + \sum_i \sum_t \text{Var}(w(\log q_t(X_t^{(i)} | X_{<t}^{(i)})))}} \\ & = \frac{\sum_i \sum_t [w(\log q_t(\tilde{X}_t^{(i)} | \tilde{X}_{<t}^{(i)}))] - \sum_i \sum_t [w(\log q_t(X_t^{(i)} | X_{<t}^{(i)})]}{\sqrt{\sum_i \text{Var}_{\tilde{X}_t \sim q_t} (\sum_t w(\log q_t(\tilde{X}_t | \tilde{X}_{<t}^{(i)}))) + \sum_i \text{Var}(\sum_t w(\log q_t(X_t^{(i)} | X_{<t}^{(i)})))}} \end{aligned}$$

For ease notations, we denoted  $\log q(X_t^{(i)} | X_{<t}^{(i)})$  as  $z_{it}^{(h)}$ , which is the logit of the  $t$ -th token of the  $i$ -th human text. Similarly, for the logit computed from machine-generated text, we define it as  $z_{it}^{(m)}$ . Using these notations, we get the objective function below:

$$\frac{1}{\sqrt{\text{Var}(\sum_t w(z_t^{(h)})) + \text{Var}(\sum_t w(z_t^{(m)}))}} \left( \sum_{i=1}^n \frac{1}{L} \sum_{t=1}^L w(z_{it}^{(h)}) - \sum_{i=1}^n \frac{1}{L} \sum_{t=1}^L w(z_{it}^{(m)}) \right). \quad (1)$$

For simplicity, (1) implicitly assumes (i) dataset includes  $n$  passages, and (ii) each passage has  $L$  tokens. Actually, the computational procedure below can be easily extended to various sample sizes and various tokens number cases.

41 Recall that the witness function have the form  $w(z) = \phi(z)^\top \beta$ , where  $\phi(z)$  are some B-spline basis  
 42 (De Boor, 1978) and  $\beta$  be the parameter. Plug in the linear function to (1), the numerator becomes:

$$\sum_{i=1}^n \frac{1}{L} \sum_{t=1}^L \phi(z_{it}^{(h)})^\top \beta - \sum_{i=1}^n \frac{1}{L} \sum_{t=1}^L \phi(z_{it}^{(m)})^\top \beta$$

43 As for the denominator, we estimate  $\text{Var}(\sum_t w(z_t^{(h)}))$  by  $\beta^\top \hat{\Sigma}^{(h)} \beta$ , and estimate  $\text{Var}(\sum_t w(z_t^{(m)}))$   
 44 by  $\beta^\top \hat{\Sigma}^{(m)} \beta$ , where in particular  $\hat{\Sigma}^{(h)} = \sum_{i=1}^n \hat{\Sigma}_i^{(h)}$  with

$$\begin{aligned} \hat{\Sigma}_i^{(h)} &= \frac{1}{L} (\mathbf{Z}_i^{(h)})^\top \mathbf{Z}_i^{(h)} - \hat{\mu}_i^{(h)} (\hat{\mu}_i^{(h)})^\top, \\ \mathbf{Z}_i^{(h)} &= \left( \phi(z_{i1}^{(h)}), \dots, \phi(z_{iL}^{(h)}) \right)^\top, \\ \hat{\mu}_i^{(h)} &= \frac{1}{L} \sum_{t=1}^L \phi(z_{it}^{(h)})^\top, \end{aligned}$$

45 and  $\hat{\Sigma}^{(m)} = \sum_{i=1}^n \hat{\Sigma}_i^{(m)}$ ; the expression for the  $\hat{\Sigma}_i^{(m)}$ 's is analogous to that of the  $\hat{\Sigma}_i^{(h)}$ 's and is  
 46 therefore omitted. Consequently, the objective function can be rewritten as:

$$\begin{aligned} & \frac{\left[ \sum_{i=1}^n \frac{1}{L} \sum_{t=1}^L \phi(z_{it}^{(h)})^\top - \sum_{i=1}^n \frac{1}{L} \sum_{t=1}^L \phi(z_{it}^{(m)})^\top \right] \beta}{\sqrt{\beta^\top (\hat{\Sigma}^{(h)} + \hat{\Sigma}^{(m)}) \beta}} \\ &= \left[ \sum_{i=1}^n \frac{1}{L} \sum_{t=1}^L \phi(z_{it}^{(h)})^\top - \sum_{i=1}^n \frac{1}{L} \sum_{t=1}^L \phi(z_{it}^{(m)})^\top \right] \beta \times \frac{1}{\|(\hat{\Sigma}^{(h)} + \hat{\Sigma}^{(m)})^{1/2} \beta\|_2} \\ &= \left[ \sum_{i=1}^n \frac{1}{L} \sum_{t=1}^L \phi(z_{it}^{(h)})^\top - \sum_{i=1}^n \frac{1}{L} \sum_{t=1}^L \phi(z_{it}^{(m)})^\top \right] \times (\hat{\Sigma}^{(h)} + \hat{\Sigma}^{(m)})^{-\frac{1}{2}} \alpha, \end{aligned}$$

47 where  $\alpha = (\hat{\Sigma}^{(h)} + \hat{\Sigma}^{(m)})^{1/2} \beta / \|(\hat{\Sigma}^{(h)} + \hat{\Sigma}^{(m)})^{1/2} \beta\|_2$  and thus  $\|\alpha\|_2 = 1$ . The maximizer of the  
 48 objective function has a closed-form solution:

$$\hat{\alpha} = \frac{\tilde{\alpha}}{\|\tilde{\alpha}\|_2}$$

49 with

$$\tilde{\alpha} = (\hat{\Sigma}^{(h)} + \hat{\Sigma}^{(m)})^{-\frac{1}{2}} \left[ \sum_{i=1}^n \frac{1}{L} \sum_{t=1}^L \phi(z_{it}^{(h)}) - \sum_{i=1}^n \frac{1}{L} \sum_{t=1}^L \phi(z_{it}^{(m)}) \right]; \quad (2)$$

50 and thus  $\|\hat{\alpha}\|_2 = 1$ . For  $\alpha$  satisfying  $\|\alpha\|_2 = 1$ , we have  $\beta = (\hat{\Sigma}^{(h)} + \hat{\Sigma}^{(m)})^{-1/2} \alpha$ . Therefore, the  
 51 expression of  $\hat{\beta}$  is:

$$\hat{\beta} = (\hat{\Sigma}^{(h)} + \hat{\Sigma}^{(m)})^{-1/2} \hat{\alpha}. \quad (3)$$

52 and  $\hat{w}(z) = \phi(z)^\top \hat{\beta}$ .

53 From Section F.3, considering the increase of n\_base would raise higher computational cost, and  
 54 thus, we recommend to set n\_base=16 and order=2.

## 55 C Assumptions for theories

56 In this section, we list the assumptions required for the theorems presented in Section 3 to hold, and  
 57 discuss when they can reasonably be expected to hold in practice and how they may be relaxed.

## 58 C.1 Assumptions

59 We work under the following assumptions on the data generating process.

60 **Assumption 1** (Margin). *With  $T_w(\bullet)$  defined as in (4) and  $w^*(\bullet)$  defined as the optimizer of (9), for*  
61 *any  $\alpha \in (0, 1)$  there are constants  $\delta_\alpha, C_\alpha$  depending only on  $\alpha$  such that for any  $x \leq \delta_\alpha$  it holds that*  
62  $\mathbb{P}_{\mathbf{X} \sim p}(|T_{w^*}(\mathbf{X}) - z_\alpha| \leq x) \leq C_\alpha x$ .

63 We also require the following technical conditions hold in order to obtain TNR lower bound and FNR  
64 control (Theorem 1 and Theorem 2).

65 **Assumption 2** (Minimum eigenvalue). *For each  $t = 1, \dots, L$  introduce the quantities*

$$\begin{aligned}\mu_t^{(1)} &= \mathbb{E}_{X_{<t} \sim p} \mathbb{E}_{\tilde{X}_t \sim q_t} \phi(\log q_t(\tilde{X}_t | X_{<t})), \\ \Sigma_t &= \mathbb{E}_{X_{<t} \sim p} \mathbb{E}_{\tilde{X}_t \sim q_t} \phi(\log q_t(\tilde{X}_t | X_{<t})) \phi(\log q_t(\tilde{X}_t | X_{<t}))^\top - \mu_t^{(1)} (\mu_t^{(1)})^\top.\end{aligned}$$

66 *There are absolute constant  $C > 0$  and  $\gamma > 0$  such that  $\lambda_{\min}(\Sigma_t) \geq Cd^{-\gamma}$  for all  $t$ .*

67 **Assumption 3** (Stochastic dominance). *For any witness function  $w \in \mathcal{W}$ ,*

$$\sum_t \mathbb{E}_{\tilde{X}_t \sim q_t} w(\log q_t(\tilde{X}_t | X_{<t})) - \mathbb{E}_{\tilde{X}_t \sim p_t} w(\log q_t(\tilde{X}_t | X_{<t})) \geq 0$$

68 *holds almost surely.*

69 **Assumption 4** (Equal variance). *For any non-constant witness function  $w$ , define*

$$\begin{aligned}\sigma_{q,L}^2 &:= \frac{1}{L} \sum_{t=1}^L \text{Var}_{\tilde{X}_t \sim q_t} \left( w(\log q_t(\tilde{X}_t | \tilde{X}_{<t})) \right), \\ \sigma_{p,L}^2 &:= \frac{1}{L} \sum_{t=1}^L \text{Var}_{\tilde{X}_t \sim p_t} \left( w(\log q_t(\tilde{X}_t | \tilde{X}_{<t})) \right).\end{aligned}$$

70  $\sigma_{q,L}^2, \sigma_{p,L}^2$  *are lower bounded by some constant  $\sigma_w^2 > 0$  almost surely. Moreover,  $\sigma_{q,L} - \sigma_{p,L} \rightarrow 0$*   
71 *in probability as  $L \rightarrow \infty$ .*

72 **Assumption 5.** *For any witness function  $w$ , define*

$$\begin{aligned}\bar{\sigma}_{q,L}^2 &= \frac{1}{L} \sum_{t=1}^L \text{Var}_{\mathbf{X} \sim q} \left( w(\log q_t(\tilde{X}_t | \tilde{X}_{<t})) \right), \\ \bar{\sigma}_{p,L}^2 &= \frac{1}{L} \sum_{t=1}^L \text{Var}_{\mathbf{X} \sim p} \left( w(\log q_t(\tilde{X}_t | \tilde{X}_{<t})) \right).\end{aligned}$$

73 *If  $\mathbf{X} \sim q$ , then  $\bar{\sigma}_{q,L}^2 / \sigma_{q,L}^2 \rightarrow 1$  in probability. If  $\mathbf{X} \sim p$ , then  $\bar{\sigma}_{p,L}^2 / \sigma_{p,L}^2 \rightarrow 1$  in probability.*

## 74 C.2 Discussion on assumptions

75 Assumption 1 is commonly assumed in the the classification and inference literature; see, for instance,  
76 Audibert & Tsybakov (2007); Shi et al. (2020). For Assumption 3, the witness function is trained to  
77 maximize the mean discrepancy. When  $w(x) \equiv 1$ , the mean discrepancy is exactly zero. Therefore,  
78 it is reasonable to restrict the function class to those functions that yield a non-negative mean  
79 discrepancy. Assumption 4 basically requires the conditional variance of logits be asymptotically  
80 equivalent for human-authored and machine-generated passages. This assumption is not overly  
81 restrictive, as the variance discrepancy between the two types of passages is relatively small in our  
82 dataset (see Figure 2). Assumption 5 is commonly assumed in martingale central limit theorem  
83 literature, see e.g. Hall & Heyde (2014); Bolthausen (1982b).

## 84 D Proofs

### 85 D.1 Notations

86 Throughout the proofs we will make use of the following notation. We will denote absolute constants  
87 by  $\kappa_1, \kappa_2, \dots$ . For a sequence of random variables  $\{X_n \mid n \geq 1\}$  with distribution functions

88  $\{F_{X_n} | n \geq 1\}$  and some (possibly degenerate) random variable  $Y$  with distribution function  $F_Y$   
 89 we write  $X_n \xrightarrow{p} Y$  as  $n \rightarrow \infty$  if  $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - Y| > \delta) = 0$  for all  $\delta > 0$ , and  $X_n \xrightarrow{d} Z$  if  
 90  $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_Y(x)$  at every continuity point of  $F_Y(\cdot)$ . For a vector  $x = (x_1, \dots, x_d)^\top \in \mathbb{R}^d$   
 91 we write  $\|x\|_p = (\sum_{j=1}^d x_j^p)^{1/p}$  with  $0 < p < \infty$  for its  $\ell_p$ -norm.

## 92 D.2 Preparatory results

93 **Theorem S1** (Bounded differences inequality). *Let  $\mathcal{X}$  be a measurable space. A function  $f : \mathcal{X}^n \rightarrow \mathbb{R}$*   
 94 *has the bounded difference property for some constants  $c_1, \dots, c_n$ , i.e., for each  $i = 1, \dots, n$ ,*

$$\sup_{\substack{x_1, \dots, x_n \\ x'_i \in \mathcal{X}}} |f(x_1, \dots, x_{i-1}, x_i, x_{i+1}, x_n) - f(x_1, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_n)| \leq c_i. \quad (4)$$

95 *Then, if  $X_1, \dots, X_n$  is a sequence of identically distributed random variables and (4) holds, putting*  
 96  *$Z = f(X_1, \dots, X_n)$  and  $\nu = \frac{1}{4} \sum_{i=1}^n c_i^2$  for any  $t > 0$ , it holds that*

$$\mathbb{P}(Z - \mathbb{E}(Z) > t) \leq e^{-t^2/(2\nu)}.$$

97 *Proof of Theorem S1.* See Section 2 in [Wainwright \(2019\)](#). □

98 **Theorem S2** (Martingale central limit theorem). *Let  $\{M_{n,i} \mid 1 \leq i \leq k_n, n \geq 1\}$  be a zero mean*  
 99 *square integrable martingale array with respect to the filtrations  $\{\mathcal{F}_{n,i} \mid 1 \leq i \leq k_n, n \geq 1\}$  having*  
 100 *increments  $X_{n,i} = M_{n,i} - M_{n,i-1}$ . If the following conditions hold*

101 **C1:**  $\sum_{i=1}^{k_n} \mathbb{E}[X_{n,i} \mathbf{1}_{\{|X_{n,i}| > \delta\}} \mid \mathcal{F}_{n,i-1}] \xrightarrow{p} 0$  as  $n \rightarrow \infty$  for all  $\delta > 0$

102 **C2:**  $\sum_{i=1}^{k_n} \mathbb{E}[X_{n,i}^2 \mid \mathcal{F}_{n,i-1}] \xrightarrow{p} \sigma^2$  as  $n \rightarrow \infty$

103 **C3:** the  $\sigma$ -fields are nested:  $\mathcal{F}_{n,i} \subseteq \mathcal{F}_{n+1,i}$  for  $1 \leq i \leq k_n$  and  $n \geq 1$

104 then  $M_{n,k_n} \xrightarrow{d} Z$  as  $n \rightarrow \infty$ , where  $Z \sim \mathcal{N}(0, \sigma^2)$ .

105 *Proof.* See Corollary 3.1 in [Hall & Heyde \(2014\)](#) and Theorem 2 in [Bolthausen \(1982a\)](#). □

106 **Lemma S1** (Convergence rate of martingale central limit theorem). *Let  $\mathbf{X} = (X_1, \dots, X_n)$  be*  
 107 *sequences of real valued random variables satisfying for all  $1 \leq t \leq n$*

$$\mathbb{E}\{X_t \mid X_{<t}\} = 0 \quad \text{almost surely.}$$

108 *Let  $\sigma_t^2 = \mathbb{E}\{X_t^2 \mid X_{<t}\}$  and  $\bar{\sigma}_t^2 = \mathbb{E}\{X_t^2\}$ . Denote  $s_n^2 = \sum_{t=1}^n \bar{\sigma}_t^2$ ,  $V_n^2 = \sum_{t=1}^n \sigma_t^2 / s_n^2$ . Suppose*  
 109  *$\|X_n\| \leq \gamma$  almost surely for all  $n$ ,  $s_n^2/n \rightarrow s^2$ , and  $V_n^2 \rightarrow 1$  as  $n \rightarrow \infty$ , then*

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P}\left(\frac{\sum_{t=1}^n X_t}{\sqrt{\sum_{t=1}^n \sigma_t^2}} \leq z\right) - \Phi(z) \right| \leq O\left(\frac{\log n}{\sqrt{n} s^3} + (\mathbb{E}|V_n^2 - 1|)^{1/3}\right),$$

110 where  $\Phi(\bullet)$  is the cumulative distribution function of standard normal distribution.

111 *Proof.* From Corollary 1 in [Bolthausen \(1982a\)](#), we have

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P}\left(\frac{\sum_{t=1}^n X_t}{s_n} \leq z\right) - \Phi(z) \right| \leq O\left(\frac{\log n}{\sqrt{n} s^3} + (\mathbb{E}|V_n^2 - 1|)^{1/3}\right).$$

112 It follows that

$$\begin{aligned}
& \sup_{z \in \mathbb{R}} \left| \mathbb{P} \left( \frac{\sum_{t=1}^n X_t}{\sqrt{\sum_{t=1}^n \sigma_t^2}} \leq z \right) - \Phi(z) \right| \\
&= \sup_{z \in \mathbb{R}} \left| \mathbb{P} \left( \frac{\sum_{t=1}^n X_t}{s_n} \leq z + \left( \frac{\sum_{t=1}^n X_t}{\sqrt{\sum_{t=1}^n \sigma_t^2}} \right) (V_n - 1) \right) - \Phi(z) \right| \\
&\leq \sup_{z \in \mathbb{R}} \mathbb{E} \left| \mathbb{P} \left( \frac{\sum_{t=1}^n X_t}{s_n} \leq z + \left( \frac{\sum_{t=1}^n X_t}{\sqrt{\sum_{t=1}^n \sigma_t^2}} \right) (V_n - 1) \right) - \Phi \left( z + \left( \frac{\sum_{t=1}^n X_t}{\sqrt{\sum_{t=1}^n \sigma_t^2}} \right) (V_n - 1) \right) \right| \\
&\quad + \sup_{z \in \mathbb{R}} \mathbb{E} \left| \Phi \left( z + \left( \frac{\sum_{t=1}^n X_t}{\sqrt{\sum_{t=1}^n \sigma_t^2}} \right) (V_n - 1) \right) - \Phi(z) \right| \\
&\leq O \left( \frac{\log n}{\sqrt{n} s^3} + (\mathbb{E}|V_n^2 - 1|)^{1/3} \right) + \sup_{z \in \mathbb{R}} |\nabla \Phi(z)| \times \mathbb{E} \left| \left( \frac{\sum_{t=1}^n X_t}{\sqrt{\sum_{t=1}^n \sigma_t^2}} \right) (V_n - 1) \right|. \tag{5}
\end{aligned}$$

113 Using the definition  $\Phi(z)$ , we know that

$$\sup_{z \in \mathbb{R}} |\nabla \Phi(z)| = \sup_{z \in \mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp(-z^2/2) \leq \frac{1}{\sqrt{2\pi}}, \tag{6}$$

114 and using the fact that  $\|X_t\|_\infty \leq \gamma$ ,  $s_n^2/n \rightarrow s^2 > 0$  and  $V_n \rightarrow 1$  in probability, we obtain that

$$\begin{aligned}
\mathbb{E} \left| \left( \frac{\sum_{t=1}^n X_t}{\sqrt{\sum_{t=1}^n \sigma_t^2}} \right) (V_n - 1) \right| &= \mathbb{E} \left| \left( \frac{\sum_{t=1}^n X_t}{s_n} \right) V_n (V_n - 1) \right| \\
&\leq \left( \mathbb{E} \left\{ \frac{(\sum_{t=1}^n X_t)^2}{s_n^2} V_n^2 \right\} \mathbb{E} \{(V_n - 1)^2\} \right)^{1/2} \\
&= O((\mathbb{E}|V_n^2 - 1|)^{1/2}), \tag{7}
\end{aligned}$$

115 where the inequality is Cauchy inequality. Combining equations (5), (6) and (7), we obtain

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P} \left( \frac{\sum_{t=1}^n X_t}{\sqrt{\sum_{t=1}^n \sigma_t^2}} \leq z \right) - \Phi(z) \right| \leq O \left( \frac{\log n}{\sqrt{n} s^3} + (\mathbb{E}|V_n^2 - 1|)^{1/3} \right), \tag{8}$$

116 which finishes the proof.  $\square$

117 **Lemma S2.** Suppose  $X$  is a random variable. Let  $\Phi(\bullet)$  be the cumulative distribution function of  
118 standard normal distribution and  $\Phi'(\bullet)$  be its derivative. Then for any random variable  $X$ ,

$$\mathbb{E}\Phi(z_\alpha + X) \geq \min\{1 - \alpha, \alpha + \Phi'(z_\alpha)\mathbb{E}X\},$$

119 where  $0 < \alpha < 1/2$ ,  $z_\alpha$  is the  $\alpha$ -th quantile of standard normal distribution.

120 *Proof of Lemma S2.* Since  $\phi(x) = (\sqrt{2\pi})^{-1} \exp(-x^2/2)$ , we noted that  $\phi(x) \geq \phi(z_\alpha)$  holds if  
121 and only if  $z_\alpha \leq x \leq z_{1-\alpha}$ . Therefore, if  $0 \leq X < z_{1-\alpha} - z_\alpha$ , then by the mean value theorem,

$$\Phi(z_\alpha + X) = \Phi(z_\alpha) + \Phi'(\xi)X \geq \alpha + \Phi'(z_\alpha)X,$$

122 where  $\xi$  lies between  $z_\alpha$  and  $z_{1-\alpha}$ . If  $X \leq 0$ , then

$$\Phi(z_\alpha + X) = \Phi(z_\alpha) + \Phi'(\eta)X \geq \alpha + \Phi'(z_\alpha)X,$$

123 where  $\eta$  lies between  $X$  and  $z_\alpha$ . Moreover, if  $X \geq z_{1-\alpha} - z_\alpha$ , then  $z_\alpha + X \geq z_{1-\alpha}$ . It follows that  
124  $\Phi(z_\alpha + X) \geq \Phi(z_{1-\alpha}) = 1 - \alpha$ . Therefore,

$$\begin{aligned}
\mathbb{E}\Phi(z_\alpha + X) &\geq \mathbb{E} \min \{ \alpha + \Phi'(z_\alpha)X, 1 - \alpha \} \\
&\geq \min \{ \alpha + \Phi'(z_\alpha)\mathbb{E}X, 1 - \alpha \},
\end{aligned}$$

125 where the last inequality follows from Jensen's inequality. This finishes the proof.  $\square$

126 The following Lemma is essential to derive a sharper lower bound for Theorem 1.

127 **Lemma S3.** Suppose  $X$  is a random variable which satisfies  $X \geq 0$  almost surely. Let  $\Phi(\bullet)$  be the  
 128 cumulative distribution function of standard normal distribution and  $\phi(\bullet)$  be the density function of  
 129 standard normal distribution. Then for any random variable  $X$ ,

$$\mathbb{E}\Phi(z_\alpha + X) \geq \sup_{\beta \in (0, \alpha]} \min\{1 - \beta, \alpha + \Phi'(z_\beta)\mathbb{E}X\},$$

130 where  $0 < \alpha < 1/2$ ,  $z_\alpha$  is the  $\alpha$ -th quantile of standard normal distribution.

131 *Proof of Lemma S3.* Since  $\Phi'(x) = (\sqrt{2\pi})^{-1} \exp(-x^2/2)$ , we noted that if  $0 < \alpha < 1/2$ , then  
 132 for any  $\beta \in (0, \alpha]$ ,  $\Phi'(x) \geq \Phi'(z_\beta)$  holds if and only if  $z_\beta \leq x \leq z_{1-\beta}$ . Therefore, if  $0 \leq X <$   
 133  $z_{1-\beta} - z_\alpha$ , then by mean value theorem,

$$\Phi(z_\alpha + X) = \Phi(z_\alpha) + \Phi'(\xi)X \geq \alpha + \Phi'(z_\beta)X,$$

134 where  $\xi$  lies between  $z_\alpha$  and  $z_{1-\beta}$ . Moreover, if  $X \geq z_{1-\beta} - z_\alpha$ , then  $z_\alpha + X \geq z_{1-\beta}$ . It follows  
 135 that  $\Phi(z_\alpha + X) \geq \Phi(z_{1-\beta}) = 1 - \beta$ . Therefore,

$$\begin{aligned} \mathbb{E}\Phi(z_\alpha + X) &\geq \mathbb{E} \min\{\alpha + \Phi'(z_\beta)X, 1 - \beta\} \\ &\geq \min\{\alpha + \Phi'(z_\beta)\mathbb{E}X, 1 - \beta\}, \end{aligned}$$

136 where the last inequality follows from Jensen's inequality. Take supremum with respect to  $\beta \in (0, \alpha]$   
 137 on both sides, we obtain

$$\mathbb{E}\Phi(z_\alpha + X) \geq \sup_{\beta \in (0, \alpha]} \min\{1 - \beta, \alpha + \Phi'(z_\beta)\mathbb{E}X\}.$$

138 This finishes the proof. Noted that by taking  $\beta = \alpha$ , the conclusion corresponds to Lemma S2.  $\square$

139 In Lemma S4 below, we provide a bound for the parameter estimation. Before doing so, we introduce  
 140 some related functions

$$Q_*(\beta) = \{\beta^\top \Sigma \beta\}^{-\frac{1}{2}} \beta^\top \mu \quad (9)$$

$$\widehat{Q}_n(\beta) = \{\beta^\top \widehat{\Sigma}_n \beta\}^{-\frac{1}{2}} \beta^\top \widehat{\mu}_n, \quad n \in \mathbb{N} \quad (10)$$

141 where in particular we have put  $\widehat{\mu}_n = L^{-1} \sum_{t=1}^L \widehat{\mu}_t^{(1)} - \widehat{\mu}_t^{(2)}$  and  $\mu = L^{-1} \sum_{t=1}^L \mu_t^{(1)} - \mu_t^{(2)}$  with

$$\widehat{\mu}_t^{(1)} = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\widetilde{X}_t \sim q_t} \phi \left( \log q_t \left( \widetilde{X}_t \mid X_{<t}^{(i)} \right) \right)$$

$$\widehat{\mu}_t^{(2)} = \frac{1}{n} \sum_{i=1}^n \phi \left( \log q_t \left( X_t^{(i)} \mid X_{<t}^{(i)} \right) \right)$$

$$\mu_t^{(1)} = \mathbb{E}_{X_{<t} \sim p} \mathbb{E}_{X_t \sim q_t} \phi \left( \log q_t \left( X_t \mid X_{<t} \right) \right)$$

$$\mu_t^{(2)} = \mathbb{E}_{X_{<t} \sim p} \mathbb{E}_{X_t \sim p_t} \phi \left( \log q_t \left( X_t \mid X_{<t} \right) \right)$$

142 for each  $t = 1, \dots, L$ , as well as  $\widehat{\Sigma}_n = L^{-1} \sum_{t=1}^L \widehat{\Sigma}_t$  and  $\Sigma = L^{-1} \sum_{t=1}^L \Sigma_t$  with

$$\widehat{\Sigma}_t = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\widetilde{X}_t \sim q_t} \left[ \phi \left( \log q_t \left( \widetilde{X}_t \mid X_{<t}^{(i)} \right) \right) \phi \left( \log q_t \left( \widetilde{X}_t \mid X_{<t}^{(i)} \right) \right)^\top \right] - \widehat{\mu}_t^{(1)} \left( \widehat{\mu}_t^{(1)} \right)^\top$$

143 for each  $t = 1, \dots, L$ .

144 **Lemma S4.** Grant the assumptions in Section C hold. Let  $\beta^*$  be the maximizer of the function (9) over  
 145 all  $\beta$ 's with  $\ell_2$  norm equal to 1 and let  $\widehat{\beta}$  be the maximizer of the empirical counterpart (10). There are  
 146 absolute constants  $\kappa_1$  and  $\kappa_2$  depending only on the constants stated in the assumptions such that for  
 147 any  $z > 0$  it holds that  $\|\widehat{\beta} - \beta^*\|_2 \leq z$  with probability at least  $1 - \kappa_1 \exp(-\kappa_2 d^{-5\gamma} n (\min\{z, 1\})^2)$ .  
 148

149 *Proof.* Observing the  $\hat{\beta} \in \arg \max_{\beta} \hat{Q}_n(\beta)$  and  $\beta^* \in \arg \max_{\beta} Q^*(\beta)$  for any  $z > 0$

$$\begin{aligned} \mathbb{P}\left(\|\hat{\beta} - \beta^*\|_2 > z\right) &= \mathbb{P}\left(\sup_{\beta: \|\beta - \beta^*\|_2 > z} Q_n(\beta) - Q_n(\beta^*) > 0\right) \\ &\leq \mathbb{P}\left(\sup_{\beta: \|\beta - \beta^*\|_2 > z} |Q_n(\beta) - Q_*(\beta)| > \frac{1}{2} \inf_{\beta: \|\beta - \beta^*\|_2 > z} [Q_*(\beta^*) - Q_*(\beta)]\right) \\ &\quad + \mathbb{P}\left(|Q_n(\beta^*) - Q_*(\beta^*)| > \frac{1}{2} \inf_{\beta: \|\beta - \beta^*\|_2 > z} [Q_*(\beta^*) - Q_*(\beta)]\right). \end{aligned} \quad (11)$$

150 For  $\beta$  such that  $\|\beta - \beta^*\|_2 > z$ , due to Assumption 2 with some absolute  $\kappa_1$  we must have that

$$Q_*(\beta^*) - Q_*(\beta) \geq \kappa_1 d^{-\gamma} \|\beta - \beta^*\|_2 \geq \kappa_1 z d^{-\gamma}. \quad (12)$$

151 For any  $\beta \in \mathbb{R}^d$  with  $\|\beta\|_2 = 1$  it holds that

$$\begin{aligned} &|Q_n(\beta) - Q_*(\beta)| \\ &\leq (\beta^\top \Sigma \beta)^{-\frac{1}{2}} |\beta^\top (\hat{\mu}_n - \mu)| + |\beta^\top \hat{\mu}_n| \left| (\beta^\top \hat{\Sigma}_n \beta)^{-\frac{1}{2}} - (\beta^\top \Sigma \beta)^{-\frac{1}{2}} \right| \\ &\leq (\beta^\top \Sigma \beta)^{-\frac{1}{2}} |\beta^\top (\hat{\mu}_n - \mu)| \\ &\quad + \frac{1}{2} \left\{ \min(\beta^\top \hat{\Sigma}_n \beta, \beta^\top \Sigma \beta) \right\}^{-\frac{3}{2}} |\beta^\top \hat{\mu}_n| |\beta^\top \hat{\Sigma}_n \beta - \beta^\top \Sigma \beta| \end{aligned} \quad (13a)$$

$$\leq \{\lambda_{\min}(\Sigma)\}^{-\frac{1}{2}} |\beta^\top (\hat{\mu}_n - \mu)| \quad (13b)$$

$$+ \frac{1}{2} \left\{ \min(\lambda_{\min}(\hat{\Sigma}_n), \lambda_{\min}(\Sigma)) \right\}^{-\frac{3}{2}} |\beta^\top \hat{\mu}_n| |\beta^\top \hat{\Sigma}_n \beta - \beta^\top \Sigma \beta|, \quad (13c)$$

152 where in particular (13a) holds due to the inequality

$$\left| \frac{1}{\sqrt{x}} - \frac{1}{\sqrt{y}} \right| = \left| \frac{x - y}{\sqrt{xy}(\sqrt{x} + \sqrt{y})} \right| \leq \frac{1}{2(\min(x, y))^{\frac{3}{2}}} |x - y| \quad (14)$$

153 for  $x, y > 0$ . For (13b) note that due to the boundedness of  $\phi(\bullet)$  for any  $\beta$  with  $\|\beta\|_2 = 1$  it holds that  
 154  $\beta^\top (\hat{\mu}_n - \mu)$  is the average of  $n$  i.i.d. random variables each bounded in absolute value by a constant  
 155 which does not depend on  $\beta$ . Therefore lower bounding  $\lambda_{\min}(\Sigma) \geq \kappa_2 d^{-\gamma}$  for some absolute  $\kappa_2$   
 156 using again the boundedness of  $\phi(\bullet)$  and applying Hoeffding's inequality, for any  $z > 0$

$$\mathbb{P}\left((13b) > \frac{1}{2} \kappa_1 z d^{-\gamma}\right) \leq \mathbb{P}\left(|\beta^\top (\hat{\mu}_n - \mu)| > \kappa_3 z d^{-\frac{3\gamma}{2}}\right) \leq 2 \exp\left(-\frac{\kappa_4 n z^2}{d^{3\gamma}}\right) \quad (15)$$

157 for certain absolute  $\kappa_3, \kappa_4$ . The following argument will be valid on the event

$$\lambda_{\min}(\hat{\Sigma}_n) \geq \frac{1}{2} \lambda_{\min}(\Sigma). \quad (16)$$

158 For (13c) notice first that with  $\|\beta\|_2 = 1$  the quantity  $|\beta^\top \hat{\mu}_n|$  is almost surely bounded from above  
 159 by some absolute constant independent of  $\beta$  and  $d$ . Moreover due to the boundedness of  $\phi(\bullet)$  it is  
 160 easy to see that the statistic  $|\beta^\top \hat{\Sigma}_n \beta - \beta^\top \Sigma \beta|$  is a self-bounding function of  $n$  random variables  
 161 with constants (see equation (4))  $c_i \propto \frac{1}{n}$  for  $i = 1, \dots, n$  which again do not depend on  $\beta$ . Therefore,  
 162 on the event (16) applying Theorem S1 we obtain that

$$\mathbb{P}\left((13c) > \frac{1}{2} \kappa_1 d^{-\gamma}\right) \leq \mathbb{P}\left(|\beta^\top \hat{\Sigma}_n \beta - \beta^\top \Sigma \beta| > \kappa_5 z d^{-\frac{5\gamma}{2}}\right) \leq \exp\left(-\frac{\kappa_6 n z^2}{d^{5\gamma}}\right) \quad (17)$$

163 for certain absolute  $\kappa_5, \kappa_6$ . Finally, note that

$$\lambda_{\min}(\hat{\Sigma}) = \min_{\beta: \|\beta\|_2=1} \beta^\top \hat{\Sigma}_n \beta \geq \lambda_{\min}(\Sigma) - \max_{\beta: \|\beta\|_2=1} \beta^\top (\hat{\Sigma}_n - \Sigma) \beta \quad (18)$$

164 and arguing as in (17), the final term (18) is no larger than  $\frac{1}{2} \kappa_2 d^{-\gamma}$  with probability at least

$$1 - 2 \exp\left(-\frac{\kappa_7 n}{d^{2\gamma}}\right).$$

165 Since by Assumption 2 we must have that  $\kappa_2 d^{-\gamma} \leq \lambda_{\min}(\Sigma)$  the event (16) must hold with the above  
 166 probability. Since the above arguments hold for any  $\beta$  with  $\|\beta\|_2 = 1$ , plugging (15) and (17) back  
 167 into (11) and accounting for the event (18) the stated result follows.  $\square$



168 **D.3 Proof of Theorem 1**

169 According to the decomposition  $T_w(\mathbf{X}) = T_w^{(1)}(\mathbf{X}) - T_w^{(2)}(\mathbf{X})$  with  $T_w^{(1)}(\mathbf{X}), T_w^{(2)}(\mathbf{X})$  defined by

$$\begin{aligned} T_w^{(1)}(\mathbf{X}) &= \frac{\sum_t [w(\log q_t(X_t|X_{<t})) - \mathbb{E}_{\tilde{X}_t \sim p_t} w(\log q_t(\tilde{X}_t|X_{<t}))]}{\sqrt{\sum_t \text{Var}_{\tilde{X}_t \sim p_t}(w(\log q_t(\tilde{X}_t|X_{<t})))}} \\ T_w^{(2)}(\mathbf{X}) &= \frac{\sum_t [\mathbb{E}_{\tilde{X}_t \sim q_t} w(\log q_t(\tilde{X}_t|X_{<t})) - \mathbb{E}_{\tilde{X}_t \sim p_t} w(\log q_t(\tilde{X}_t|X_{<t}))]}{\sqrt{\sum_t \text{Var}_{\tilde{X}_t \sim q_t}(w(\log q_t(\tilde{X}_t|X_{<t})))}}, \end{aligned} \quad (19)$$

170 we obtain that the TNR can be represented as

$$\mathbb{P}_{\mathbf{X} \sim p}(T_w(\mathbf{X}) \leq z_\alpha) = \mathbb{P}_{\mathbf{X} \sim p}(T_w^{(1)}(\mathbf{X}) \leq z_\alpha + T_w^{(2)}(\mathbf{X})) \quad (20)$$

171 It is easy to verify that when  $\mathbf{X} \sim p$ ,  $T_w^{(1)}(\mathbf{X})\sigma_{q,L}/\sigma_{p,L}$  converges to standard normal distribution,  
172 using the convergence rate for martingale central limit theorem (i.e., Theorem S1), we obtain that

$$\begin{aligned} \mathbb{P}_{\mathbf{X} \sim p}(T_w(\mathbf{X}) \leq z_\alpha) &= \mathbb{P}_{\mathbf{X} \sim p}\left(T_w^{(1)}(\mathbf{X})\frac{\sigma_{q,L}}{\sigma_{p,L}} \leq (z_\alpha + T_w^{(2)}(\mathbf{X}))\frac{\sigma_{q,L}}{\sigma_{p,L}}\right) \\ &\geq \Phi(z_\alpha + T_w^{(2)}(\mathbf{X})) + \left(\Phi\left((z_\alpha + T_w^{(2)}(\mathbf{X}))\frac{\sigma_{q,L}}{\sigma_{p,L}}\right) - \Phi(z_\alpha + T_w^{(2)}(\mathbf{X}))\right) \\ &\quad + O\left(\log L/\sqrt{L}\right) \\ &\geq \Phi(z_\alpha + T_w^{(2)}(\mathbf{X})) - \sup_{z \in \mathbb{R}} \Phi'(z) \times |z_\alpha + T_w^{(2)}(\mathbf{X})| \times \left|\frac{\sigma_{q,L}}{\sigma_{p,L}} - 1\right| \\ &\quad + O\left(\log L/\sqrt{L}\right). \end{aligned}$$

173 Take expectation on both sides, we obtain

$$\mathbb{P}_{\mathbf{X} \sim p}(T_w(\mathbf{X}) \leq z_\alpha) \geq \mathbb{E}\Phi(z_\alpha + T_w^{(2)}(\mathbf{X})) + o(1) + O(\log L/\sqrt{L}).$$

174 Now define  $\tilde{\sigma}_{q,L}^2 = \mathbb{E}_{\mathbf{X} \sim p}\sigma_{q,L}^2$ . Then under equal variance assumption (Assumption 4), we also have  
175  $\sigma_{q,L} - \tilde{\sigma}_{q,L} \rightarrow 0$  in probability. It follows that for any  $\epsilon > 0$ ,

$$\begin{aligned} \mathbb{E}\Phi(z_\alpha + T_w^{(2)}(\mathbf{X})) &= \mathbb{E}\Phi(z_\alpha + T_w^{(2)}(\mathbf{X}))\mathbb{I}\{|\sigma_{q,L} - \tilde{\sigma}_{q,L}| \leq \epsilon\} \\ &\quad + \mathbb{E}\Phi(z_\alpha + T_w^{(2)}(\mathbf{X}))\mathbb{I}\{|\sigma_{q,L} - \tilde{\sigma}_{q,L}| > \epsilon\} \\ &\geq \mathbb{E}\Phi(z_\alpha + T_w^{(2)}(\mathbf{X}))\mathbb{I}\{|\sigma_{q,L} - \tilde{\sigma}_{q,L}| \leq \epsilon\} \\ &= \mathbb{E}\Phi\left(z_\alpha + T_w^{(2)}(\mathbf{X})\frac{\sigma_{q,L}}{\tilde{\sigma}_{q,L} - \epsilon}\right)\mathbb{I}\{|\sigma_{q,L} - \tilde{\sigma}_{q,L}| \leq \epsilon\} \\ &\geq \mathbb{E}\Phi\left(z_\alpha + T_w^{(2)}(\mathbf{X})\frac{\sigma_{q,L}}{\tilde{\sigma}_{q,L} - \epsilon}\right) \\ &\quad - \mathbb{E}\Phi\left((z_\alpha + T_w^{(2)}(\mathbf{X}))\frac{\sigma_{q,L}}{\tilde{\sigma}_{q,L} - \epsilon}\right)\mathbb{I}\{|\sigma_{q,L} - \tilde{\sigma}_{q,L}| > \epsilon\} \\ &\geq \mathbb{E}\Phi\left(z_\alpha + T_w^{(2)}(\mathbf{X})\frac{\sigma_{q,L}}{\tilde{\sigma}_{q,L} - \epsilon}\right) - \mathbb{P}(|\sigma_{q,L} - \tilde{\sigma}_{q,L}| > \epsilon), \end{aligned}$$

176 where the first inequality is obtained due to  $\Phi$  is non-negative and the last inequality holds because  
177  $T_w^{(2)}(\mathbf{X})\sigma_{q,L} > 0$  under stochastic dominance assumption (Assumption 3) and  $\Phi$  is bounded by 1.  
178 Together with Lemma S2 and Assumption 4, we obtain

$$\begin{aligned} &\mathbb{P}_{\mathbf{X} \sim p}(T_w(\mathbf{X}) \leq z_\alpha) \\ &\geq \min\left\{1 - \alpha, \alpha + \phi(z_\alpha)\mathbb{E}\left\{T_w^{(2)}(\mathbf{X})\frac{\sigma_{q,L}}{\tilde{\sigma}_{q,L}}\right\}\right\}\frac{\tilde{\sigma}_{q,L}}{\tilde{\sigma}_{q,L} - \epsilon} \\ &\quad - \mathbb{P}\{|\sigma_{q,L} - \tilde{\sigma}_{q,L}| \geq \epsilon\} + O\left(\log L/\sqrt{L}\right) + o(1). \end{aligned} \quad (21)$$

Let  $L \rightarrow \infty$  and using the fact that  $\mathbb{E} \left\{ T_w^{(2)}(\mathbf{X})^{\frac{\sigma_{q,L}}{\sigma_{q,L}}} \right\} = T_w^{(2*)}(\mathbf{X})$ , we obtain that TNR is asymptotically lower bounded by  $\min\{1 - \alpha, \alpha + \phi(z_\alpha) T_w^{(2*)}(\mathbf{X})\}^{\frac{\tilde{\sigma}_{q,L}}{\sigma_{q,L} - \epsilon}}$ . By taking  $\epsilon \rightarrow 0$ , then the conclusion of Theorem 1 follows.

**Remark 1.** It is worth noting that since  $T_w^{(2)}(\mathbf{X})\sigma_{q,L} > 0$ , a sharper lower bound can be obtained by applying Lemma S3 in the last step (inequality (21)), then follow a similar argument, we will obtain that TNR is asymptotically lower bounded by

$$\sup_{0 < \beta \leq \alpha} \min\{1 - \beta, \alpha + \phi(z_\beta) T_w^{(2*)}(\mathbf{X})\}. \quad (22)$$

Noted that with any fixed  $\alpha \in (0, 1/2)$ , the lower bound in (22) may tend to 1 given that  $T_w^{(2*)}(\mathbf{X})$  is sufficiently large, which is a sharper bound than in Theorem 1.

#### D.4 Proof of Theorem 2

*Proof.* Denote  $Z_t = \hat{w}(\log q_t(X_t|X_{<t})) - \mathbb{E}_{\tilde{X}_t \sim q_t(\bullet|X_{<t})} \hat{w}(\log q_t(X_t|X_{<t}))$ . Then if  $\mathbf{X} \sim q$ , we have  $\mathbb{E}\{Z_t|X_{<t}\} = 0$  almost surely. Under Assumptions 3, 5 and the boundedness of  $\hat{w}$  (the B-spline basis are bounded), it is easy to verify that  $Z_t$  satisfies all conditions of Lemma S1. Therefore, according to Lemma S1, we obtain for any  $\alpha \in (0, 1)$ ,

$$\begin{aligned} \text{FNR}_{\hat{w}} - \alpha &= \mathbb{P}_{\mathbf{X} \sim q}(T_{\hat{w}}(\mathbf{X}) \leq z_\alpha) - \Phi(z_\alpha) \\ &= \mathbb{P}_{\mathbf{X} \sim q} \left( \frac{\sum_{t=1}^L Z_t}{\sum_{t=1}^L \mathbb{E}\{Z_t^2|X_{<t}\}} \leq z_\alpha \right) - \Phi(z_\alpha) \\ &\leq O\left(\frac{\log L}{\sqrt{L}}\right) + O((\mathbb{E}|V_L - 1|)^{1/3}). \end{aligned}$$

Taking expectation on both sides, we obtain

$$\mathbb{E}(\text{FNR}_{\hat{w}}) \leq \alpha + O\left(\frac{\log L}{\sqrt{L}}\right) + O((\mathbb{E}|V_L - 1|)^{1/3}).$$

This completes the proof.  $\square$

#### D.5 Proof of Theorem 3

*Proof.* Since  $\mathbb{E}(\text{TNR}_{\hat{w}}) \geq \text{TNR}_{w^*} - \mathbb{E}(|\text{TNR}_{\hat{w}} - \text{TNR}_{w^*}|)$ , it is enough to upper bound the second term in the last expression. Denote by  $\hat{T}_n(\bullet)$  and  $\hat{T}^*(\bullet)$  respectively the classifier (4) using witness functions  $\hat{w}(\bullet) = \phi(\bullet)^\top \hat{\beta}$  and  $w^*(\bullet) = \phi(\bullet)^\top \beta^*$  (see the definition of  $\hat{\beta}$  and  $\beta^*$  in Lemma S4). Write  $\hat{\Delta}_n(\mathbf{x}) = |\hat{T}_n(\mathbf{x}) - T^*(\mathbf{x})|$  and  $\hat{\Delta}_n = \sup_{\mathbf{x}} \Delta_n(\mathbf{x})$ . For any  $z_\alpha > 0$  we have that

$$\begin{aligned} |\text{TNR}_{\hat{w}} - \text{TNR}_{w^*}| &= \left| \mathbb{P}_{\mathbf{X} \sim p}(\hat{T}_n(\mathbf{X}) \leq z_\alpha) - \mathbb{P}_{\mathbf{X} \sim p}(T^*(\mathbf{X}) \leq z_\alpha) \right| \\ &= \left| \int \mathbf{1}_{\{\hat{T}_n(\mathbf{x}) \leq z_\alpha\}} - \mathbf{1}_{\{T^*(\mathbf{x}) \leq z_\alpha\}} d\mathbf{p}(\mathbf{x}) \right| \\ &\leq \int \left| \mathbf{1}_{\{\hat{T}_n(\mathbf{x}) \leq z_\alpha\}} - \mathbf{1}_{\{T^*(\mathbf{x}) \leq z_\alpha\}} \right| d\mathbf{p}(\mathbf{x}) \\ &= \int \mathbf{1}_{\{\hat{T}_n(\mathbf{x}) \leq z_\alpha, T^*(\mathbf{x}) > z_\alpha\}} + \mathbf{1}_{\{\hat{T}_n(\mathbf{x}) > z_\alpha, T^*(\mathbf{x}) \leq z_\alpha\}} d\mathbf{p}(\mathbf{x}) \\ &\leq 2 \int \mathbf{1}_{\{|T^*(\mathbf{x}) - z_\alpha| \leq \hat{\Delta}_n\}} d\mathbf{p}(\mathbf{x}) \\ &= 2\mathbb{P}_{\mathbf{X} \sim p}(|T^*(\mathbf{X}) - z_\alpha| \leq \hat{\Delta}_n). \end{aligned} \quad (23)$$

Due to Assumption 1 on the event

$$\{\hat{\Delta}_n \leq \delta_0\} \quad (24)$$

we will have that  $|\text{TNR}_{\hat{w}} - \text{TNR}_{w^*}| \leq \kappa_3 \hat{\Delta}_n$  for some absolute  $\kappa_3$ . We therefore focus on bounding the quantity  $\hat{\Delta}_n$ . For each  $w \in \Omega$  and each  $j = 1, \dots, L$ , we introduce the quantities:

$$\begin{aligned} Y_j^{(w)} &= w(\log q_t(X_t | X_{<t})), \\ \mu_j^{(w)} &= \mathbb{E}_{\tilde{X}_t \sim q_t} w(\log q_t(\tilde{X}_t | X_{<t})), \\ (\sigma_j^{(w)})^2 &= \text{Var}_{\tilde{X}_t \sim q_t} w(\log q_t(\tilde{X}_t | X_{<t})). \end{aligned}$$

With this notation in place we have that for any  $\mathbf{x}$

$$\hat{\Delta}_n(\mathbf{x}) \leq \frac{1}{\sqrt{L}} \left| \sum_{j=1}^L y_j^{(\hat{w})} - \mu_j^{(\hat{w})} \right| \left| \sqrt{L^{-1} \sum_{j=1}^L (\sigma_j^{(\hat{w})})^2} - \sqrt{L^{-1} \sum_{j=1}^L (\sigma_j^{(w^*)})^2} \right|^{-1} \quad (25a)$$

$$+ \left\{ \frac{1}{L} \sum_{t=1}^L (\sigma_j^{(w^*)})^2 \right\}^{-\frac{1}{2}} \frac{1}{\sqrt{L}} \left| \sum_{j=1}^L (Y_j^{(\hat{w})} - Y_j^{(w^*)}) - (\mu_j^{(\hat{w})} - \mu_j^{(w^*)}) \right|, \quad (25b)$$

where for clarity we have suppressed dependence on  $\mathbf{x}$  above. For ease of notation put  $Z_t = \log q_t(X_t | X_{<t})$  and  $\tilde{Z}_t = \log q_t(\tilde{X}_t | X_{<t})$  where  $\tilde{X}_t \sim q_t$ . Write also  $\phi(\bullet) = (B_1(\bullet), \dots, B_d(\bullet))^\top$ . Recalling that  $w(\bullet) = \phi(\bullet)^\top \beta$  for arbitrary  $j = 1, \dots, L$  we have

$$\begin{aligned} \left| (\sigma_j^{(\hat{w})})^2 - (\sigma_j^{(w^*)})^2 \right| &\leq \mathbb{E} \left[ \sum_{l_1=1}^d \sum_{l_2=1}^d \left| \hat{\beta}_{l_1} \hat{\beta}_{l_2} - \beta_{l_1}^* \beta_{l_2}^* \right| (|B_{l_1}(z_j) B_{l_2}(z_j)| \right. \\ &\quad \left. + \left| \mathbb{E}[B_{l_1}(\tilde{Z}_j)] \mathbb{E}[B_{l_2}(\tilde{Z}_j)] \right| + 2 |B_{l_1}(z_j) \mathbb{E}[B_{l_2}(\tilde{Z}_j)]| \right) ] \\ &\leq \frac{\kappa_4}{2} \sum_{l_1=1}^d \sum_{l_2=1}^d \left| \hat{\beta}_{l_1} \hat{\beta}_{l_2} - \beta_{l_1}^* \beta_{l_2}^* \right| \\ &= \frac{\kappa_4}{2} \sum_{l_1=1}^d \sum_{l_2=1}^d \left| \hat{\beta}_{l_1} (\hat{\beta}_{l_2} - \beta_{l_2}^*) - \beta_{l_2}^* (\hat{\beta}_{l_1} - \beta_{l_1}^*) \right| \\ &\leq \frac{\kappa_4}{2} \sum_{l_1=1}^d \sum_{l_2=1}^d \left| \hat{\beta}_{l_1} \right| \left| \hat{\beta}_{l_2} - \beta_{l_2}^* \right| + \frac{\kappa_4}{2} \sum_{l_1=1}^d \sum_{l_2=1}^d \left| \beta_{l_2}^* \right| \left| \hat{\beta}_{l_1} - \beta_{l_1}^* \right| \\ &= \kappa_5 \sqrt{d} \left\| \hat{\beta} - \beta^* \right\|_1 \kappa_5 \\ &\leq d \left\| \hat{\beta} - \beta^* \right\|_2, \end{aligned}$$

for absolute  $\kappa_4, \kappa_5$ . Consequently, using inequality (14) and Assumption 2, on the event (18) we obtain that with absolute  $\kappa_6$ :

$$(25a) \leq \kappa_6 d^3 \frac{1}{\sqrt{L}} \left| \sum_{j=1}^L y_j^{(\hat{w})} - \mu_j^{(\hat{w})} \right| \left\| \hat{\beta} - \beta^* \right\|_2. \quad (26)$$

Observe that conditional on  $\hat{\beta}$  the term  $\sum_{j=1}^L (y_j^{(\hat{w})} - \mu_j^{(\hat{w})})$  is a martingale with increments bounded from above almost surely by a constant independent on  $\hat{\beta}$ ; by the Azuma–Hoeffding inequality the normalized sum in (26) has sub-Gaussian tails. By Lemma S4 the term  $\left\| \hat{\beta} - \beta^* \right\|_2$  likewise has sub-Gaussian tails. Therefore on the relevant events we obtain that (26) has sub-exponential tails, and consequently for any  $z > 0$

$$\mathbb{P}(26 > z) \leq \kappa_7 \exp \left( -\kappa_8 \min \left\{ z^2 \frac{n}{d^{5\gamma+6}}, z \sqrt{\frac{n}{d^{5\gamma+6}}} \right\} \right) \quad (27)$$

for certain absolute  $\kappa_7, \kappa_8$ . Similar arguments show that the normalized sum in (25b) has the same tail behavior as (27). Since the above arguments do not depend on  $\mathbf{x}$  we obtain that (27) likewise

described the tail behavior of  $\widehat{\Delta}_n$ . Consequently, on the relevant events we obtain that

$$\begin{aligned}
\mathbb{E} |\text{TNR}_{\widehat{w}} - \text{TNR}_{w^*}| &= \int_0^\infty \mathbb{P}(|\text{TNR}_{\widehat{w}} - \text{TNR}_{w^*}| > z) \, dz \\
&\leq \int_0^\infty \mathbb{P}(\kappa_3 \widehat{\Delta}_n > z) \, dz \\
&\leq \kappa_9 \int_0^\infty \exp\left(-\kappa_{10} \min\left\{z^2 \frac{n}{d^{5\gamma+6}}, z\sqrt{\frac{n}{d^{5\gamma+6}}}\right\}\right) \, dz \\
&\leq \kappa_{11} \sqrt{\frac{d^{5\gamma+6}}{n}}
\end{aligned} \tag{28}$$

for certain absolute  $\kappa_9, \kappa_{10}, \kappa_{11}$ . When the events (24) and (14) do not hold from (23) we have the conservative bound  $\mathbb{E} |\text{TNR}_{\widehat{w}} - \text{TNR}_{w^*}| \leq 2$ . However, the probability of these events not holding is smaller than (28) up to constants. Therefore, the stated result follows by the law of total expectation.  $\square$

## E Experiment details

**Pre-trained language models.** We assess the performance of our method using text generated from various pre-trained language models outlined in Table S1. Following the setting in Bao et al. (2024), for the models with over 6B parameters, we employ half-precision (`torch.float16`), otherwise, we use full-precision (`torch.float32`).

Name	Model <sup>†</sup>	Scale (Billion)
GPT-2 (Radford et al., 2019)	openai-community/gpt2-xl	1.5B
GPT-Neo (Black et al., 2021)	EleutherAI/gpt-neo-2.7B	2.7B
OPT-2.7 (Zhang et al., 2022)	facebook/opt-2.7b	2.7B
GPT-J (Wang & Komatsuzaki, 2021)	EleutherAI/gpt-j-6B	6B
GPT-NeoX (Black et al., 2022)	EleutherAI/gpt-neox-20b	20B

Table S1: Description of the source models that is used to produce machine-generated text. <sup>†</sup>: we present the address of models in <https://huggingface.co/>.

**Implementations of baselines.** For the baselines considered in our experiments, we use the existing implementation provided in <https://github.com/baoguangsheng/fast-detect-gpt>, which is distributed in the MIT License. We run DetectGPT and NPR with default 100 perturbations with the T5 model (Raffel et al., 2020) and run DNA-GPT with a truncate-ratio of 0.5 and 10 prefix completions per passage.

**Evaluation Metric.** We measure the detection accuracy by AUC (short for “area under the curve”). AUC ranges from 0.0 to 1.0, an AUC of 1.0 indicates a perfect classifier and vice versa. The relative improvement of AdaDetectGPT over FastDetectGPT is calculated by  $\frac{\text{AdaDetectGPT} - \text{FastDetectGPT}}{1.0 - \text{FastDetectGPT}}$ , which represents how much improvement has been made relative to the maximum possible improvement for FastDetectGPT.

**Hardware details.** Most of experiments are run on a Tesla A100 GPU (40GB) with 10 vCPU Intel Xeon Processor and 72GB RAM. For the experiments where the source model is GPT-NeoX, we run on a H20-NVLink (96GB) GPU with 20 vCPU Intel(R) Xeon(R) Platinum and 200GB RAM.

## F Additional results

### F.1 Distribution of statistics on open-source models

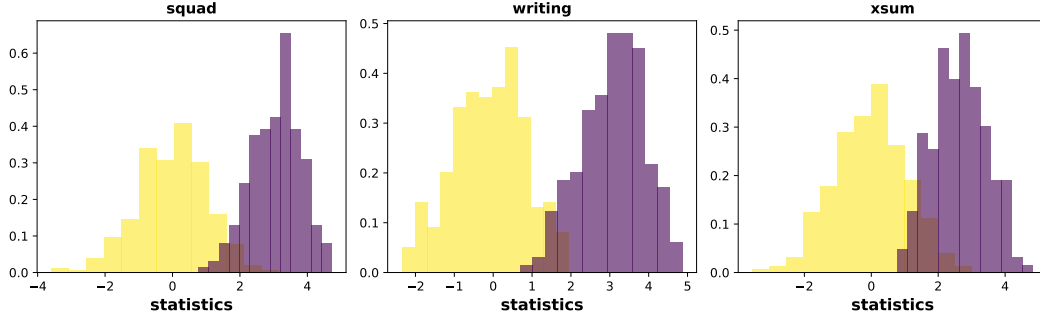


Figure S1: Histogram of statistics in three dataset. Each panel visualizes the histogram of statistics. The yellow histogram corresponds to the case when the sampled texts exactly follow the conditional probability of the source model, while purple histogram corresponds to text drawn with deviations from that distribution.

### F.2 Accuracy on open-source models under black-box setting

Dataset	Method	Source Model				Avg.
		GPT-2	OPT-2.7	GPT-Neo	GPT-NeoX	
SQuAD	FastDetectGPT	0.6181	0.6495	0.6230	0.6910	0.6813
	AdaDetectGPT	0.6920	0.7195	0.7382	0.7338	0.7460
	Relative	19.3570	19.9651	30.5609	13.8495	20.2957
	FastDetectGPT <sup>†</sup>	0.8145	0.8166	0.9220	0.7519	0.8188
	AdaDetectGPT <sup>†</sup>	<b>0.8249</b>	<b>0.8308</b>	<b>0.9273</b>	<b>0.7609</b>	<b>0.8300</b>
	Relative	5.6301	7.7245	6.7968	3.6121	6.2106
Writing	FastDetectGPT	0.7662	0.7918	0.7685	0.8022	0.8028
	AdaDetectGPT	0.8306	0.8529	0.8555	<b>0.8587</b>	0.8636
	Relative	27.5699	29.3365	37.6112	28.5350	30.8124
	FastDetectGPT <sup>†</sup>	0.8565	0.8497	0.9215	0.8182	0.8582
	AdaDetectGPT <sup>†</sup>	<b>0.8780</b>	<b>0.8737</b>	<b>0.9386</b>	0.8567	<b>0.8849</b>
	Relative	14.9666	15.9741	21.7742	21.1856	18.8023
XSum	FastDetectGPT	0.5919	0.6445	0.5718	0.6389	0.6468
	AdaDetectGPT	0.6795	0.7238	0.6879	0.7045	0.7261
	Relative	21.4569	22.2991	27.1129	18.1580	22.4439
	FastDetectGPT <sup>†</sup>	0.8145	0.8166	0.9220	0.7519	0.8188
	AdaDetectGPT <sup>†</sup>	<b>0.8249</b>	<b>0.8308</b>	<b>0.9273</b>	<b>0.7609</b>	<b>0.8300</b>
	Relative	9.8060	10.5637	10.2543	8.1057	11.1574

Table S2: Zero-shot detection accuracy on five source models under the black-box setting. <sup>†</sup>: use two surrogate models for sampling and scoring, where the sampling model is GPT-J while the scoring model is GPT-Neo.

### F.3 Analysis factors in training $w$

Since AdaDetectGPT requires training a witness function, we systematically examine three key factors influencing its performance: (i) the size of the training set; (ii) tuning parameters for generating B-spline basis and (iii) distribution shift between training and test data.

**Robust performance on various training data sizes.** We evaluate AdaDetectGPT across varying dataset sizes by setting  $n_1 = n_2 \in \{100, 200, 300, 400, 500, 600\}$  for both human- and machine-generated texts. Figure S2 demonstrates that AdaDetectGPT has a clear performance advantage over

FastDetectGPT when sample size is large. This is expected because a larger sample size leads to a more accurate estimation of  $w$ . Notably, even with limited data  $n_1 = n_2 = 100$ , AdaDetectGPT maintains superior accuracy compared to baseline methods, though the performance gap decreases. These results highlight our method’s effectiveness on learning the witness function.

**Insensitivity on tuning parameters.** The two critical tuning parameters of B-splines are: (i) the number of basis functions ( $n_{\text{base}}$ ) and (ii) the maximum polynomial order. Our experiments fix one parameter while varying the other (with  $n_{\text{base}}=16$  or  $\text{order}=2$  as defaults). As shown in Figure S3 in Appendix, AdaDetectGPT achieves the highest AUC scores so long as  $n_{\text{base}} \geq 4$ . Besides, enlarging  $n_{\text{base}}$  improves the AUC of AdaDetectGPT although the improvement becomes marginal when  $n_{\text{base}} \geq 16$ . Figure S3 also shows that increasing the polynomial order from linear to quadratic visibly improve the performance; while increasing order from quadratic to cubic/quartic has a limited gain. Having said that, even when the B-splines function is piecewise linear, our method still outperform all baselines.

**Robust against distribution shift.** In this part, we create training sets with different distributions than the test data by varying the number of human prompt tokens in machine-generated text. In contrary, for the test data, the number of human prompt tokens are fixed. As shown in Figure S4, AdaDetectGPT demonstrates high robustness - the distributional discrepancy between training and test data has negligible effect on classification accuracy. Notably, our method always achieves the highest AUC across all experimental setup.

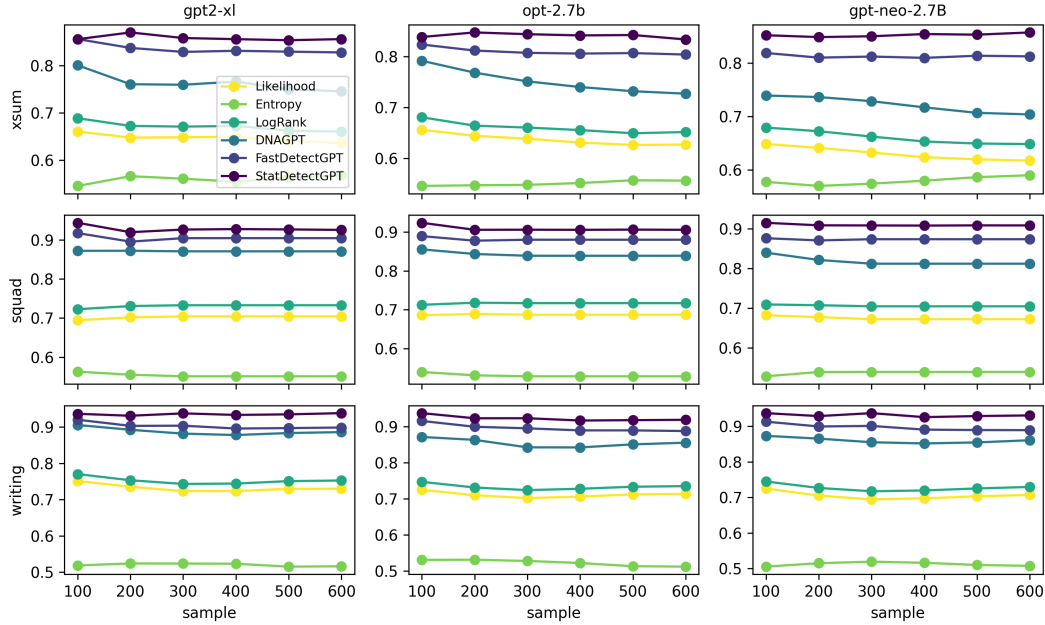


Figure S2: Classification accuracy versus the sample size for training  $w$ . We omit DetectGPT, NPR, and DNA in this experiments as they are time-consuming.

## G Broader impact and limitation

AdaDetectGPT is a statistically efficient detector for machine-generated text. By accurately detecting generated passages, it can help safeguard AI systems against fake news, disinformation, and academic plagiarism.

In this paper, we prove rigorous guarantees in the white-box setting and demonstrate empirical performance in the black-box setting. The promising results under black-box setting motivate future work to establish matching theoretical guarantees. Moreover, even in the white-box setting, practical LLM’s text generation often employs sampling parameters (e.g. `temperature` and `top_k`) that makes the sampling distribution differs from the conditional distribution derived from the source model. This mismatch causes MCLT may not hold in practice. Fortunately, we observe that our test statistic still shifts toward a positive mean (see Figure S1), implying FNR remains control.

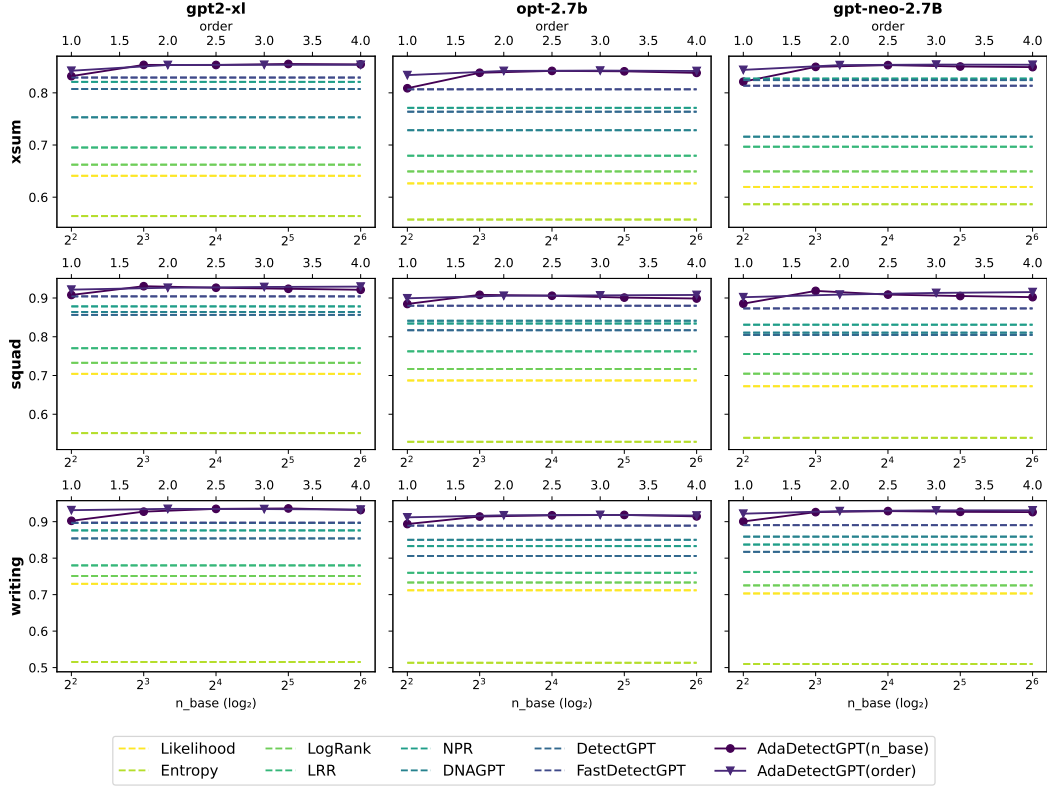


Figure S3: The classification accuracy of AdaDetectGPT and baseline methods. AdaDetectGPT( $n_{\text{base}}$ ) present the AUC when the number of basis in B-spline increases as 4, 8, 16, 32, 64 (bottom  $x$ -axis); while AdaDetectGPT( $\text{order}$ ) shows the AUC when the maximum order of basis in B-spline increases from 1 to 4 (top  $x$ -axis). The AUC of baseline methods are presented by dash lines.

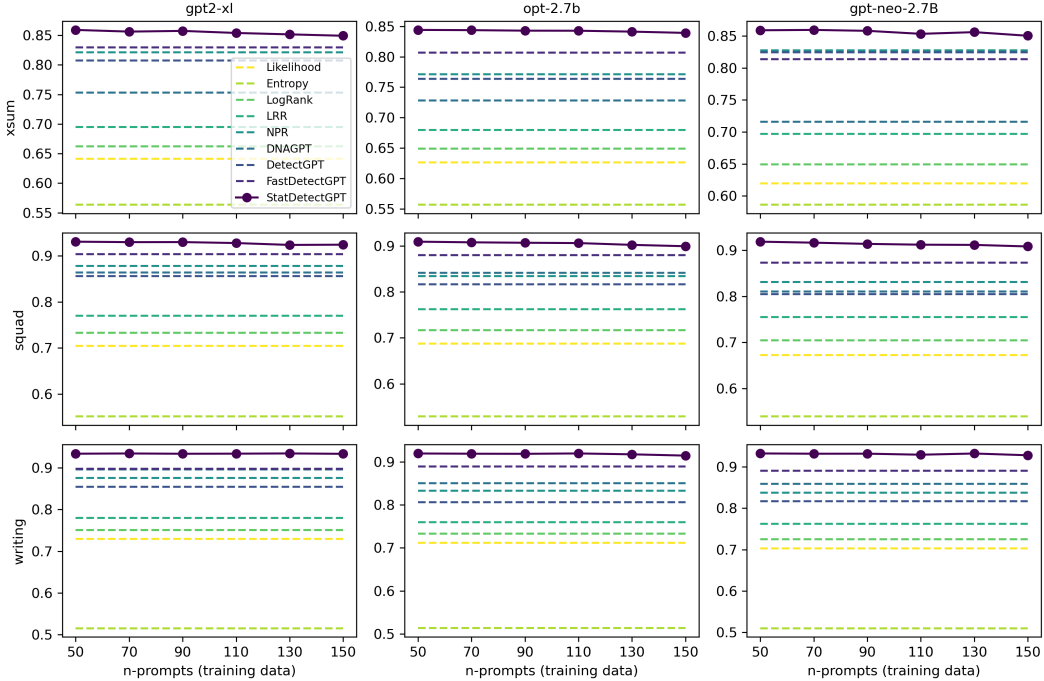


Figure S4: The classification accuracy of AdaDetectGPT when the number of human prompts changes. The AUC of baseline methods are presented by dash lines.

## References

- Audibert, J.-Y. and Tsybakov, A. B. Fast learning rates for plug-in classifiers. *Annals of Statistics*, 35 (2):608–633, 2007.
- Bao, G., Zhao, Y., Teng, Z., Yang, L., and Zhang, Y. Fast-detectGPT: Efficient zero-shot detection of machine-generated text via conditional probability curvature. In *The Twelfth International Conference on Learning Representations*, 2024. URL <https://openreview.net/forum?id=Bpcgcr8E8Z>.
- Black, S., Leo, G., Wang, P., Leahy, C., and Biderman, S. GPT-Neo: Large Scale Autoregressive Language Modeling with Mesh-Tensorflow, March 2021. URL <https://doi.org/10.5281/zenodo.5297715>. If you use this software, please cite it using these metadata.
- Black, S., Biderman, S., Hallahan, E., Anthony, Q., Gao, L., Golding, L., He, H., Leahy, C., McDonnell, K., Phang, J., Pieler, M., Prashanth, U. S., Purohit, S., Reynolds, L., Tow, J., Wang, B., and Weinbach, S. Gpt-neox-20b: An open-source autoregressive language model, 2022. URL <https://arxiv.org/abs/2204.06745>.
- Bolthausen, E. Exact Convergence Rates in Some Martingale Central Limit Theorems. *The Annals of Probability*, 10(3):672 – 688, 1982a. doi: 10.1214/aop/1176993776. URL <https://doi.org/10.1214/aop/1176993776>.
- Bolthausen, E. Exact Convergence Rates in Some Martingale Central Limit Theorems. *The Annals of Probability*, 10(3):672 – 688, 1982b. doi: 10.1214/aop/1176993776. URL <https://doi.org/10.1214/aop/1176993776>.
- De Boor, C. *A practical guide to splines*, volume 27. springer New York, 1978.
- Hall, P. and Heyde, C. C. *Martingale limit theory and its application*. Academic press, 2014.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67, 2020. URL <http://jmlr.org/papers/v21/20-074.html>.
- Shi, C., Lu, W., and Song, R. Breaking the curse of nonregularity with subagging—inference of the mean outcome under optimal treatment regimes. *Journal of Machine Learning Research*, 21(176): 1–67, 2020.
- Wainwright, M. J. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press, 2019.
- Wang, B. and Komatsuzaki, A. GPT-J-6B: A 6 Billion Parameter Autoregressive Language Model. <https://github.com/kingoflolz/mesh-transformer-jax>, May 2021.
- Zhang, S., Roller, S., Goyal, N., Artetxe, M., Chen, M., Chen, S., Dewan, C., Diab, M., Li, X., Lin, X. V., Mihaylov, T., Ott, M., Shleifer, S., Shuster, K., Simig, D., Koura, P. S., Sridhar, A., Wang, T., and Zettlemoyer, L. Opt: Open pre-trained transformer language models, 2022.