Check for updates

Adv. Appl. Probab. **58,** 475–525 (2026)
*doi:*10.1017/apr.2025.10029

ENHANCED UPPER CONFIDENCE LIMITS VIA RANDOMIZED TESTS IN RANDOM SAMPLING WITHOUT REPLACEMENT

ZIHAO LI ,* ** AND
HUANGJUN ZHU ,* *** *Fudan University*
MASAHITO HAYASHI ,**** *The Chinese University of Hong Kong*

Abstract

In this paper we study one-sided hypothesis testing under random sampling without replacement, which frequently appears in the cryptographic problem setting, including the verification of measurement-based quantum computation. Suppose that $n+1$ binary random variables $X_1, \ldots, X_{n+1}$ follow a permutation invariant distribution and n binary random variables $X_1, \ldots, X_n$ are observed. Then, we propose randomized tests with a randomization parameter for the expectation of the $(n+1)$th random variable X_{n+1} under a given significance level $\delta > 0$. Our randomized tests significantly improve the upper confidence limit over deterministic tests. Our problem setting commonly appears in machine learning in addition to cryptographic scenarios by considering adversarial examples. Such studies are essential for expanding the applicable area of statistics. Although this paper addresses only binary random variables, a similar significant improvement by randomized tests can be expected for general non-binary random variables.

Keywords: Randomized test; random sampling without replacement; adversarial scenario; adversarial example

2010 Mathematics Subject Classification: Primary 62F05; 62D05
Secondary 62F03

1. Introduction

Random sampling without replacement is prevalent across various scientific disciplines, including social sciences, computer sciences, and biology. Its applications span a wide range, from modeling gene mutations [30] and recommender systems [4] to analyzing social networks [8] and implementing variational clustering algorithms [43]. In addition, beyond these traditional applications, random sampling can be employed for learning in cryptographic scenarios in which adversaries may exist. In cryptographic settings, when an adversarial player generates

Received 22 July 2025; accepted 22 July 2025.
* Postal address: Department of Physics and State Key Laboratory of Surface Physics, Institute for Nanoelectronic Devices and Quantum Computing, Center for Field Theory and Particle Physics, Fudan University, Shanghai 200433, China.
** Email address: zihaoli20@fudan.edu.cn
*** Email address: zhuhuangjun@fudan.edu.cn
**** Postal address: School of Data Science, The Chinese University of Hong Kong, Shenzhen, Longgang District, Shenzhen 518172, China; International Quantum Academy (SIQA), Futian District, Shenzhen 518048, China; and Graduate School of Mathematics, Nagoya University, Chikusa-ku, Nagoya 464-8602, Japan. Email: hmasahito@cuhk.edu.cn

© The Author(s), 2025. Published by Cambridge University Press on behalf of Applied Probability Trust.

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

random variables [41], these variables can be arbitrarily correlated. In contrast, as explained in Section 2, when a standard player generates random variables, they are subjected to the independent and identically distributed (i.i.d.) condition. Such cryptographic settings encompass numerous applications in quantum information processing that demand high-security requirements, such as blind-measurement-based quantum computation [22] and quantum key distribution [23]. Moreover, similar settings are observed in machine learning, referred to as adversarial examples [7, 9, 15, 20, 28, 33]. In such settings, to estimate the behavior of the random variables, we need to employ random sampling without replacement. In addition, this kind of problem setting has been actively studied in the area of quantum information in several references (see [46, 47] and references therein), while its essential part can be treated as a problem of statistics like the problem presented in this paper.

Random sampling without replacement is described by the hypergeometric distribution. It models the number of samples per group given a limited total number of samples. Since the selection of a single group element affects the probability of the remaining elements being drawn, the behavior of the random variable differs from that under the i.i.d. condition. Hypothesis testing provides a more logically complete structure to decision-making and therefore to better decisions. However, most existing studies on statistical hypothesis testing assume the i.i.d. condition, making them unsuitable for the adversarial setting mentioned previously. Although many papers have studied the behavior of observations under random sampling without replacement [1, 2, 5, 10–12, 14, 16, 18, 19, 24, 32, 35, 36, 38, 40, 42, 45], only a limited number of works [3, 39] have addressed the problem under the framework of hypothesis testing. Therefore, to establish a logical decision-making framework in the above areas, it is necessary to study hypothesis testing under the hypergeometric distribution. In fact, while this setting is useful in the cryptographic setting, it can also be applied to other scenarios beyond the cryptographic setting that involve adversarial examples [7, 9, 15, 20, 28, 33]. In these scenarios, the adversarial case can be considered as a certain accidental case, and the i.i.d. case is considered as a certain non-accidental case.

In this paper we discuss one-sided hypothesis testing for random sampling without replacement. Assume that $n+1$ binary random variables $X_1, \ldots, X_{n+1}$ are generated according to random sampling without replacement. The task of this paper is to verify whether the expectation of the $(n+1)$th binary random variable X_{n+1} is smaller than a threshold ϵ when n observations $X_1, \ldots, X_n$ are given. We investigate this problem under the framework of statistical hypothesis testing, i.e. we study the upper confidence limit for the above expectation under a given significance level $\delta > 0$. In the case of random sampling with replacement, the optimal test has been well studied. In the optimal test, the decision should be made based on whether the count K of observed outcome 0 among $X_1, \ldots, X_n$ is greater than or less than a specific threshold $k(n, \delta)$ [27]. Only when K is equal to this threshold is a randomized decision necessary. This indicates that the benefit gained from employing a randomized decision is marginal.

Suppose that we have $n+m$ random variables $X_1, \ldots, X_{n+m}$ and n observations, and we are interested in the expectation of the remaining m variables. If both n and m go to infinity, we can use the normal approximation, which enables us to handle our problem in a way similar to the case with normal distribution. However, when random variables $X_1, \ldots, X_{n+1}$ are generated and we have only one remaining variable, we cannot use a normal approximation. Hypothesis testing in this case has not been sufficiently studied. This case requires us to discuss the problem in a different way. Such a problem setting occurs very frequently when we consider a cryptographic setting. In particular, the authors' companion paper [29] applies the

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

results obtained in this paper to the verification of blind quantum computation based on the relation pointed in [22], although quantum key distribution is related to the case when n and m increase simultaneously [23]. In fact, the recent paper [29] derived the analysis of the verification of blind quantum computation based on the result of this paper. In addition, it is expected that the above-mentioned problem setting for hypothesis testing appears in adversarial machine learning [41].

In this paper, to address the problem mentioned above, we introduce a randomized test that makes a randomized decision using a randomization parameter $0 \leq \lambda < 1$ when the outcome K is greater than a certain threshold l. That is, our randomized test makes a deterministic decision when the outcome K is not greater than the threshold l. This randomized test with parameter $0 < \lambda < 1$ significantly improves the performance over the deterministic test. For example, when the significance level $\delta > 0$ is exponentially small in n, the upper confidence limit for the expectation of X_{n+1} is always the trivial value 1, i.e. any deterministic test cannot provide a non-trivial value. However, our randomized test provides a non-trivial value for the upper confidence limit even with an exponentially small significance level δ in n. We study the asymptotic behavior of the performance of this new randomized test under two regimes, depending on the behavior of the threshold l. One is the constant regime in which the threshold l remains a constant. The other is the linear regime in which the threshold l increases in proportion to n. Under both regimes, we show that our randomized tests outperform the deterministic test significantly.

The rest of this paper is organized as follows. Section 2 explains the relation between our problem setting and the cryptographic setting. Section 3 formulates our problem setting and introduces our new randomized test based on the randomization parameter $0 \leq \lambda < 1$. Some preliminary results that demonstrate the randomized test's advantages are also presented in Section 3, with detailed proofs deferred to Section 4. Section 5 presents the first main result (Theorem 1) of this paper, which clarifies the asymptotic behaviors of our randomized test under the linear regime. Section 6 considers the power, i.e. the probability of detecting a given upper confidence limit for the i.i.d. case under this regime. Section 7 presents the second main result (Theorem 2) of this paper, which clarifies the asymptotic behaviors of our randomized test under the constant regime. Section 8 proposes a new method for computing the upper confidence limit, with its analytical expression provided in Theorem 3. This result is pivotal to the proofs of our main results in Sections 5 and 7. Section 9 presents a brief proof of the theorem stated in Section 5. Section 10 gives the conclusion. Several auxiliary lemmas are presented in Appendices A and B. Many technical results are also presented in the appendices.

2. Relation with Cryptographic Setting

The research topics in cryptography are not limited to secret message transmission; message authentication is also a major focus in this field. Any message authentication process involves a decision-making step to determine whether a received document should be accepted or rejected, as we need to consider two scenarios, the honest case and the dishonest case, where the latter is often referred to as the adversarial case. Beyond message authentication, it is often required to distinguish the honest case and the dishonest case in the cryptographic setting. For example, in the verification of quantum computation [29, 31, 46, 47], there are two players, Alice and Bob, with Bob preparing part of the quantum computing device for Alice to use. Assume that Alice does not trust Bob, then she must measure the part provided by Bob to verify whether he has prepared it honestly. This kind of problem setting is formulated in the following general way.

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

In the honest case, the original random variables $X_1, X_2, \ldots, X_{n+1}$ satisfy the i.i.d. condition. In the dishonest case, these variables do not satisfy this condition. In the latter scenario, it is presumed that the adversary has artificial control over these random variables. As such, our decision-making process must account for the dishonest case. The cryptographic scenario imposes the following two requirements on our decision-making approach [31].

The first requirement is soundness, which ensures the quality of the remaining variable Y_{n+1} (whose formal definition is given in Section 3) under the condition that the test is passed and the dishonest case is assumed. Since we consider 'under the condition', we need to guarantee that the conditional probability $\mathrm{P}(Y_{n+1} = 1|\,\mathrm{Pass}\,)$ is smaller than a certain threshold. This is because we are interested in the objective probability only when the test is passed. In the case of a one-sided test, soundness is guaranteed by the upper (or lower) confidence limit of the parameter of interest under a given significant level. In other words, this part corresponds to the calculation of the minimum significant level to guarantee a given upper confidence limit, which is often called the p-value for the upper confidence limit.

The second requirement is completeness, which means that the test is passed with a high probability in the honest case when the probability $\mathrm{P}(X_i = 1)$ is sufficiently small. The completeness is guaranteed by the power. Therefore, we discuss the power of our test only in the honest case.

3. Problem Setting

3.1. Formulation

We consider $n+1$ binary random variables $X_1, \ldots, X_{n+1}$ that take values in $\{0, 1\}$ and are subject to a joint distribution $P_{X_1,\ldots,X_{n+1}}$. By randomly drawing n samples $Y_1, \ldots, Y_n$ without replacement from $X_1, \ldots, X_{n+1}$, we would like to estimate the marginal distribution of the remaining binary variable Y_{n+1}. To characterize this problem, we introduce the set S_{n+1} of permutations on $[n+1] := \{1, \ldots, n+1\}$. The joint distribution $P_{Y_1,\ldots,Y_{n+1}}$ of $Y_1, \ldots, Y_{n+1}$ is given as

$$P_{Y_1,\ldots,Y_{n+1}}(y_1, \ldots, y_{n+1}) = \sum_{\sigma \in S_{n+1}} \frac{1}{(n+1)!} P_{X_{\sigma(1)},\ldots,X_{\sigma(n+1)}}(y_1, \ldots, y_{n+1}).$$

Denote the set of permutation invariant distributions on $[n+1]$ by $\mathcal{Q}_{n+1}$, then the joint distribution $P_{Y_1,\ldots,Y_{n+1}}$ can be considered as an arbitrary element of $\mathcal{Q}_{n+1}$.

Our problem setting is formulated as follows. Suppose n observed variables $Y_1, \ldots, Y_n$ are $y_1, \ldots, y_n$, our task is to estimate the conditional distribution $P_{Y_{n+1}|(Y_1,\ldots,Y_n)=(y_1,\ldots,y_n)}$. For each observed variable Y_i, the observation outcome $Y_i = 0$ is labeled as a 'success' and the outcome $Y_i = 1$ is labeled as a 'failure'. Since $P_{Y_1,\ldots,Y_{n+1}}$ is permutation invariant, the above conditional distribution $P_{Y_{n+1}|Y_1,\ldots,Y_n}$ depends only on the random variable K that describes the number of failures among the n observations $Y_1, \ldots, Y_n$. Hence, we discuss only the conditional distribution $P_{Y_{n+1}|K}$. Since Y_{n+1} is a binary variable taking values in $\{0, 1\}$, it suffices to estimate the conditional probability $\mathrm{P}(Y_{n+1} = 1|K \leq k)$ for each realization k.

Now, we discuss this problem in the framework of one-sided hypothesis testing with significance level δ. The problem setting is formulated as follows. We are interested in the probability of $Y_{n+1} = 1$ under the condition that our test is passed. That is, we focus on the conditional distribution $\mathrm{P}(Y_{n+1} = 1|\,\mathrm{Pass}\,)$, where 'Pass' means the case that our test is passed. The aim of our test is to guarantee that the conditional probability $\mathrm{P}(Y_{n+1} = 1|\,\mathrm{Pass}\,)$ is not greater than a threshold ϵ with significance level δ. Hence, when its upper bound ϵ is given, the null

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

hypothesis H_0 and the alternative hypothesis H_1 are given as

$$H_0: \quad \mathrm{P}(Y_{n+1} = 1 | \text{ Pass }) > \epsilon,$$
$$H_1: \quad \mathrm{P}(Y_{n+1} = 1 | \text{ Pass }) \leq \epsilon.$$

In this problem setting, it is needed to set the event 'Pass' as follows: when the test is passed, the null hypothesis H_0 is rejected. This is because we are required to guarantee the alternative hypothesis H_1 under the event 'Pass'. Due to the above requirement and the correlation between Y_{n+1} and $Y_1, \ldots, Y_n$, the forms of our hypotheses depend on our decision.

In other words, the expectation $\mathrm{E}[Y_{n+1}]$ does not equal $\mathrm{P}(Y_{n+1} = 1 | \text{ Pass })$. This point is the difficulty of our problem. Given a test T, the maximum of the conditional probability $\mathrm{P}(Y_{n+1} = 1 | \text{ Pass } T)$ with significance level δ is given as

$$\overline{\epsilon}(T, n, \delta) := \max_{Q \in \mathcal{Q}_{n+1}} \{Q(Y_{n+1} = 1 | \text{ Pass } T) | Q(\text{ Pass } T) \geq \delta\}.$$

Since we can guarantee that the conditional probability $\mathrm{P}(Y_{n+1} = 1 | \text{ Pass } T)$ is not greater than the maximum, the above maximum is simply called the upper confidence limit at significance level δ. In this case, we can say that δ is the p-value of the upper bound $\overline{\epsilon}(T, n, \delta)$ of the conditional probability $\mathrm{P}(Y_{n+1} = 1 | \text{ Pass } T)$.

The most natural choice for the rejection region of the null hypothesis H_0 is the set $\{K | K \leq k\}$, i.e. the test is passed when $K \leq k$. This rejection region is determined by the threshold value k. We denote this deterministic test by T_k. Under this choice of the rejection region, the upper confidence limit is given as

$$\overline{\epsilon}(T_k, n, \delta) = \max_{Q \in \mathcal{Q}_{n+1}} \{Q(Y_{n+1} = 1 | K \leq k) \mid Q(K \leq k) \geq \delta\}.$$

That is, we can guarantee that the conditional probability $\mathrm{P}(Y_{n+1} = 1 | K \leq k)$ is not greater than the upper confidence limit $\overline{\epsilon}(T_k, n, \delta)$ at significance level δ. Here the threshold k denotes the number of allowed failures.

As we show in Section 4.1, the upper confidence limit $\overline{\epsilon}(T_k, n, \delta)$ reads

$$\overline{\epsilon}(T_k, n, \delta) = \begin{cases} \dfrac{(k+1)(n+1-k) - \delta(n+1)}{\delta(n-k)(n+1)} & \text{when } \delta \geq \dfrac{k+1}{n+1}, \\ 1 & \text{when } \delta < \dfrac{k+1}{n+1}. \end{cases} \tag{1}$$

Equation (1) shows that when the significance level δ is smaller than $1/(n+1) = \min_{k \geq 0} ((k+1)/(n+1))$, we cannot make any non-trivial upper confidence limit for the conditional probability $\mathrm{P}(Y_{n+1} = 1 | K \leq k)$ for any integer $k = 0, 1, \ldots, n$. Therefore, it is impossible to realize an exponentially small significance level with respect to n for any non-trivial upper confidence limit. In other words, when we set the significance level δ to be an exponentially small value with respect to n, the upper confidence limit becomes 1.

Fortunately, this problem can be resolved by introducing a randomized test. Our randomized test is not a trivial idea. To see the strength of our method, we first see a usual randomized test as follows. When we consider a randomized test, usually we make the randomization only on the threshold. Consider the following test. When $K \leq k-1$, we reject the null hypothesis H_0. When $K \geq k+1$, we adopt the null hypothesis H_0. When $K = k$, we reject the null hypothesis H_0 with probability $1 - \lambda$, and adopt H_0 with probability λ. This test can be considered as a

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

probabilistic mixture of the test T_k and T_{k-1} with probabilities $1-\lambda$ and λ. We denote this test by $\tilde{T}_k(\lambda)$. Since $\tilde{T}_k(0) = T_k$, this method is a generalization of the above method. As we show in Section 4.2, the upper confidence limit $\overline{\epsilon}(\tilde{T}_k(\lambda), n, \delta)$ reads

$$\overline{\epsilon}(\tilde{T}_k(\lambda), n, \delta) = \begin{cases} \xi_1(n, k, \lambda, \delta) & \text{when } \delta \ge (1-\lambda) + \lambda \dfrac{k}{n+1}, \\ \xi_2(n, k, \lambda, \delta) & \text{when } (1-\lambda) + \lambda \dfrac{k}{n+1} > \delta \ge (1-\lambda)\dfrac{k+1}{n+1}, \\ 1 & \text{when } (1-\lambda)\dfrac{k+1}{n+1} > \delta, \end{cases} \tag{2}$$

where

$$\xi_1(n, k, \lambda, \delta) := \frac{(k-1)\lambda(n+1-k) + (n+1)(1-\delta)}{\delta(n+1)\lambda(n+1-k)},$$

$$\xi_2(n, k, \lambda, \delta) := \frac{\delta(n+1)(k-(1-\lambda)(k+1)) + (1-\lambda)^2(k+1)(n+1-k)}{\delta(n+1)(k+(1-\lambda)(n-2k))}.$$

Consequently, with this type of randomized test, when we set the significance level δ to be an exponentially small value with respect to n, the upper confidence limit $\overline{\epsilon}(\tilde{T}_k(\lambda), n, \delta)$ becomes 1. To see this, we examine two cases depending the behavior of λ. First, if $(1-\lambda)(k+1)/(n+1) > \delta$ then the above observation holds immediately. Second, if $(1-\lambda) + \lambda k/(n+1) > \delta \ge (1-\lambda)(k+1)/(n+1)$ then $(1-\lambda)$ is also an exponentially small value with respect to n, leading to that $\overline{\epsilon}(\tilde{T}_k(\lambda), n, \delta) = \xi_2(n, k, \lambda, \delta)$ converges to 1. This discussion shows that a more advanced randomization is needed for further improvement.

The above randomized test makes a randomized decision only when the observation is the threshold value. In contrast, our method makes this kind of randomized decision for all observations, and the probability of this randomization depends on our observation. To introduce our randomized test, which is different from the above randomized test, we consider the following artificial noise before our data processing. Instead of $Y_1, \ldots, Y_n$, we focus on n variables $U_1, \ldots, U_n$ with the following conditional distribution $P_{U_i|Y_i}$:

$$P_{U_i|Y_i}(0|0) = 1, \quad P_{U_i|Y_i}(1|0) = 0, \quad P_{U_i|Y_i}(0|1) = \lambda, \quad P_{U_i|Y_i}(1|1) = 1-\lambda, \tag{3}$$

where $0 \le \lambda < 1$ is a randomization parameter. For convenience, we define $\nu := 1-\lambda$. The outcome $U_i = 0$ is labeled as a 'success' and the outcome $U_i = 1$ is labeled as a 'failure'. Denote the number of failures among $U_1, \ldots, U_n$ by the random variable L, which satisfies the following conditional distribution $P_{L|K}$:

$$P_{L|K}(l|k) = \begin{cases} 0 & \text{when } l > k, \\ \binom{k}{l} \lambda^{k-l}(1-\lambda)^l & \text{when } l \le k, \end{cases}$$

which implies that $L \le K$. Then, we define the following test by using the random variable L instead of K. When $L \le l$, we reject the null hypothesis H_0. When $L \ge l+1$, we adopt the null hypothesis H_0. We denote this test by $T_l(\lambda)$, in which the threshold l denotes the number of allowed failures.

Then, the upper confidence limit $\overline{\epsilon}(T_l(\lambda), n, \delta)$ is calculated as

$$\overline{\epsilon}(T_l(\lambda), n, \delta) = \max_{Q \in \mathcal{Q}_{n+1}} \{Q(Y_{n+1} = 1 | L \le l) \mid Q(L \le l) \ge \delta\}.$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

When $K \le l$, the relation $L \le K$ guarantees that the probability $\mathrm{P}(Y_{n+1} = 1)$ under the observed condition is not greater than the upper limit. Otherwise, we make a randomized decision. Note that $\overline{\epsilon}(T_l, n, \delta)$ is the special case of $\overline{\epsilon}(T_l(\lambda), n, \delta)$ with $\lambda = 0$. This method guarantees that the conditional probability $\mathrm{P}(Y_{n+1} = 1 | L \le l)$ is not greater than $\overline{\epsilon}(T_l(\lambda), n, \delta)$ at significance level δ.

The one-sided hypothesis testing introduced above plays a crucial role in the task of verifying quantum states in the adversarial scenario (QSVadv) [46, 47]. In QSVadv, one aims to ensure that the quantum state on one system has sufficiently high fidelity with the target state by performing n measurements on other systems. So far, only the verification protocol with $l = 0$ has been studied in the literature. The performance with general $l > 0$ is still not clear.

To see how our new method with $\lambda > 0$ can achieve exponentially small significance level, we set $l = 0$. For $z \in \{0, 1, \ldots, n+1\}$, let $\delta = \delta_z := ((n+1-z)\lambda^z + z\lambda^{z-1})/(n+1)$, then we have

$$\overline{\epsilon}(T_0(\lambda), n, \delta_z) = \frac{z}{z + (n - z + 1)\lambda}. \tag{4}$$

Equation (4) follows from (57) in Ref. [47], and a simple proof is also given in Section 4.3. Since δ_z has a similar behavior as λ^z, this equation shows that the above randomized test can realize a non-trivial statement even with an exponentially small significance level. So randomized tests can significantly outperform deterministic tests.

By contrast, randomized tests are not necessary to achieve an exponentially small significance level for i.i.d. variables. To see this, we define the upper confidence limit for the i.i.d. case and a test T as

$$\overline{\epsilon}_{\mathrm{i.i.d.}}(T, n, \delta) := \max_{Q \in \mathcal{Q}_{\mathrm{i.i.d.},n+1}} \{Q(Y_{n+1} = 1 | \text{ Pass } T) \mid Q(\text{ Pass } T) \ge \delta\},$$

where $\mathcal{Q}_{\mathrm{i.i.d.},n+1}$ is the set of independent and identical distributions on $[n+1]$. Then, for the deterministic test T_k, we have

$$\overline{\epsilon}_{\mathrm{i.i.d.}}(T_k, n, \delta) = \max_{Q \in \mathcal{Q}_{\mathrm{i.i.d.},n+1}} \{Q(Y_{n+1} = 1 | K \le k) \mid Q(K \le k) \ge \delta\}. \tag{5}$$

We denote by P_θ^{n+1} the independent and identical distribution on $[n+1]$ where the probability of the event $Y_i = 1$ is θ. Hence, $\mathcal{Q}_{\mathrm{i.i.d.},n+1}$ can be written as $\{P_\theta^{n+1}\}_{\theta \in [0,1]}$, which implies that

$$\overline{\epsilon}_{\mathrm{i.i.d.}}(T_k, n, \delta) = \max_{\theta \in [0,1]} \{\theta \mid P_\theta^{n+1}(K \le k) \ge \delta\}. \tag{6}$$

As we show in Sections 5 and 7, $\overline{\epsilon}_{\mathrm{i.i.d.}}(T_k, n, \delta)$ has a non-trivial value even when the significance level δ is exponentially small with respect to n.

3.2. Frequently used notation

In preparation for further discussions, here we introduce further notation, which will be frequently used throughout the paper.

Let $\mathbb{Z}^{\ge j}$ be the set of integers larger than or equal to j. For $0 \le p \le 1$ and $z, l, j \in \mathbb{Z}^{\ge 0}$, let

$$b_{z,j}(p) := \binom{z}{j} p^j (1-p)^{z-j}, \qquad B_{z,l}(p) := \sum_{j=0}^{l} b_{z,j}(p) = \sum_{j=0}^{l} \binom{z}{j} p^j (1-p)^{z-j}. \tag{7}$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

Here it is understood that $b_{z,j}(p) = 0$ whenever $j \geq z+1$, and $x^0 = 1$ even if $x = 0$. When there is no danger of confusion, we shall use $b_{z,l}$ and $B_{z,l}$ as shorthand for $b_{z,l}(\nu)$ and $B_{z,l}(\nu)$, respectively. Define

$$\Delta_{z,l} := B_{z,l} - B_{z+1,l}, \quad z, l = 0, 1, 2, \ldots. \tag{8}$$

For $0 < \delta \leq 1$, define

$$z^*(l, \delta, \lambda) := \min\{z \in \mathbb{Z}^{\geq l} | B_{z,l} \leq \delta\}, \quad z_*(l, \delta, \lambda) := z^*(l, \delta, \lambda) - 1. \tag{9}$$

Here $z^*(l, \delta, \lambda)$ and $z_*(l, \delta, \lambda)$ are well defined because $B_{l,l}(\nu) = 1$, $\lim_{z\to\infty} B_{z,l}(\nu) = 0$, and that $B_{z,l}(\nu)$ is non-increasing in z for $z \geq l$ according to Lemma 7. In this paper we shall frequently use z^* and z_* as shorthand for $z^*(l, \delta, \lambda)$ and $z_*(l, \delta, \lambda)$, respectively.

The standard normal density function and its distribution function read

$$\phi(x) := \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}, \quad \Phi(x) := \int_{-\infty}^{x} \phi(t)\, \mathrm{d}t. \tag{10}$$

The function $\Phi^{-1}(x)$ is defined as the inverse function of $\Phi(x)$.

For $0 < p, q < 1$, the relative entropy between two binary non-negative vectors $(p, 1-p)$ and $(q, 1-q)$ reads

$$D(p\|q) := p \ln \frac{p}{q} + (1-p) \ln \frac{1-p}{1-q}. \tag{11}$$

For $\gamma \geq 0$ and $0 < x < 1$, define $t_{\mathrm{D}}(\gamma, x)$ as the unique solution of t to the equation

$$tD\left(\frac{\gamma}{t} \middle\| x\right) = 1, \quad t \geq \frac{\gamma}{x}. \tag{12}$$

Note that $t_{\mathrm{D}}(\gamma, x)$ is well defined because $D(x\|x) = 0$, $D(0\|x) = -\ln(1-x)$, and that $D(\gamma/t\|x)$ is non-decreasing in t for $t \geq \gamma/x$ thanks to Lemma 12 in Appendix A.2. For $0 \leq s < 1$ and $r > 0$, define $\epsilon_D(s, r)$ as the unique solution of ϵ to the equation

$$r = D(s\|\epsilon), \quad 0 \leq s \leq \epsilon < 1. \tag{13}$$

Similarly, for $0 < \epsilon < 1$ and $0 < r \leq -\ln(1-\epsilon)$, define $s_D(\epsilon, r)$ as the unique solution of s to (13). Both $\epsilon_D(s, r)$ and $s_D(\epsilon, r)$ are well defined according to Lemma 12 in Appendix A.2.

For $x \geq 0$ and $k \in \mathbb{Z}^{\geq 0}$, the Poisson cumulative distribution function reads

$$\mathrm{Pois}(k, x) := \mathrm{e}^{-x} \sum_{j=0}^{k} \frac{x^j}{j!}. \tag{14}$$

For $0 < \delta < 1$ and $k \in \mathbb{Z}^{\geq 0}$, define $t_{\mathrm{P}}(k, \delta)$ as the unique solution of t to the equation

$$\mathrm{Pois}(k, t) = \delta \quad \text{for } t \geq 0. \tag{15}$$

Note that $t_{\mathrm{P}}(k, \delta)$ is well defined because $\mathrm{Pois}(k, 0) = 1$, $\lim_{t\to\infty} \mathrm{Pois}(k, t) = 0$, and that $\mathrm{Pois}(k, t)$ is monotonically decreasing in t for $t \geq 0$ according to Lemma 13 in Appendix A.3.

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

3.3. Properties of the upper confidence limit

Here we summarize some basic properties of the upper confidence limit. In particular, Proposition 1 clarifies the monotonicities of $\overline{\epsilon}_{\text{i.i.d.}}(T_k, n, \delta)$, and Lemma 1 provides useful bounds for $\overline{\epsilon}(T_k, n, \delta)$, $\overline{\epsilon}(T_l(\lambda), n, \delta)$, and $\overline{\epsilon}_{\text{i.i.d.}}(T_k, n, \delta)$. The monotonicities of $\overline{\epsilon}(T_l(\lambda), n, \delta)$ will be clarified later in Proposition 6 after several preparations. Proposition 1 and Lemma 1 are proved in Appendices C and D, respectively.

Proposition 1. *Suppose* $0 < \delta \le 1$, $n, k \in \mathbb{Z}^{\ge 0}$, *and* $n > k$. *Then* $\overline{\epsilon}_{\text{i.i.d.}}(T_k, n, \delta)$ *is non-increasing in* δ *for* $0 < \delta \le 1$, *non-increasing in* n *for* $n \ge k+1$, *and non-decreasing in* k *for* $0 \le k \le n-1$.

Lemma 1. *Suppose* $0 < \delta \le \frac{1}{2}$, $k, l, n \in \mathbb{Z}^{\ge 0}$, $n > k$, $n > l$. *In the i.i.d. case, we have*

$$\overline{\epsilon}_{\text{i.i.d.}}(T_k, n, \delta) \ge \overline{\epsilon}_{\text{i.i.d.}}\left(T_k, n, \frac{1}{2}\right) > \frac{k}{n}. \tag{16}$$

As a generalization, for $0 \le \lambda < 1$ *we have*

$$\overline{\epsilon}(T_l(\lambda), n, \delta) \ge \overline{\epsilon}\left(T_l(\lambda), n, \frac{1}{2}\right) \begin{cases} = 1 & \text{when } l \ge \nu n, \\ > \frac{l}{\nu n} & \text{when } l < \nu n. \end{cases} \tag{17}$$

For $\lambda = 0$ *and* $k/n < \delta \le \frac{1}{2}$, *we have*

$$\overline{\epsilon}(T_k, n, \delta) > \frac{k}{n\delta}. \tag{18}$$

4. Proofs of Simple Key Relations

4.1. Proof of (1)

Let Q_j be the permutation invariant distribution of variables $Y_1, Y_2, \ldots, Y_{n+1} \in \{0, 1\}$ that satisfy $\sum_{i=1}^{n+1} Y_i = j$. Then any element of $\mathcal{Q}_{n+1}$ can be written in the form $Q[\vec{p}] := \sum_{j=0}^{n+1} p_j Q_j$, where $\vec{p} := (p_j)_{j=0}^{n+1}$. In addition, we have

$$Q_j(K = k) = \begin{cases} \dfrac{n+1-j}{n+1} & \text{when } k = j, \\ \dfrac{j}{n+1} & \text{when } k = j-1, \\ 0 & \text{otherwise.} \end{cases}$$

It follows that

$$Q[\vec{p}](K \le k) = \sum_{j=0}^{k} p_j + \frac{k+1}{n+1} p_{k+1} \tag{19}$$

$$Q[\vec{p}](Y_{n+1} = 1, K \le k) = \sum_{j=0}^{k} \frac{j}{n+1} p_j + \frac{k+1}{n+1} p_{k+1}, \tag{20}$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

and

$$\begin{aligned}\overline{\epsilon}(T_k, n, \delta) &= \max_{(p_j)_{j=0}^{n+1}} \left\{ \frac{Q[\vec{p}](Y_{n+1}=1, K\le k)}{Q[\vec{p}](K\le k)} \,\middle|\, Q[\vec{p}](K\le k)\ge\delta \right\} \\ &\overset{(a)}{=} \max_{0\le p_k\le 1} \left\{ \frac{p_k\dfrac{k}{n+1}+\dfrac{k+1}{n+1}(1-p_k)}{p_k+\dfrac{k+1}{n+1}(1-p_k)} \,\middle|\, p_k+\frac{k+1}{n+1}(1-p_k)\ge\delta \right\} \\ &= \max_{0\le p_k\le 1} \left\{ \frac{k+1-p_k}{(k+1)+p_k(n-k)} \,\middle|\, \frac{(k+1)+p_k(n-k)}{n+1}\ge\delta \right\}. \end{aligned} \tag{21}$$

To derive (a) above, we use (19), (20), and choose $p_1, \ldots, p_{k-1}$ to be zero. When $(k+1)/(n+1) > \delta$, the maximum in (21) is realized with $p_k = 0$, i.e. the maximum is 1. When $(k+1)/(n+1) \le \delta$, the maximum is realized with $p_k = ((n+1)\delta - (k+1))/(n-k)$, i.e. the maximum is $((k+1)(n-k+1) - \delta(n+1))/[\delta(n+1)(n-k)]$. Then, we have (1).

4.2. Proof of (2)

The test $\tilde{T}_k(\lambda)$ satisfies that

$$\begin{aligned} Q[\vec{p}](\text{ Pass } \tilde{T}_j(\lambda)) &= (1-\lambda)\left(\sum_{j=0}^{k} p_j + \frac{k+1}{n+1}p_{k+1}\right) + \lambda\left(\sum_{j=0}^{k-1} p_j + \frac{k}{n+1}p_k\right) \\ &= \sum_{j=0}^{k-1} p_j + \left(1-\lambda+\frac{k}{n+1}\lambda\right)p_k + \frac{k+1}{n+1}(1-\lambda)p_{k+1} = \sum_{j=0}^{k+1} a_j p_j, \\ Q[\vec{p}](Y_{n+1}=1, \text{ Pass } \tilde{T}_j(\lambda)) &= (1-\lambda)\sum_{j=0}^{k+1}\frac{j}{n+1}p_j + \lambda\sum_{j=0}^{k}\frac{j}{n+1}p_j \\ &= \sum_{j=0}^{k}\frac{j}{n+1}p_j + \frac{k+1}{n+1}(1-\lambda)p_{k+1} = \sum_{j=0}^{k+1} b_j p_j, \end{aligned}$$

where we choose the parameters a_j and b_j as follows:

$$\begin{aligned} &a_{k+1} := (1-\lambda)\frac{k+1}{n+1}, \quad a_k := (1-\lambda)+\lambda\frac{k}{n+1}, \quad a_j := 1 \quad \text{for } j=0,\ldots,k-1, \\ &b_{k+1} := (1-\lambda)\frac{k+1}{n+1}, \quad b_j := \frac{j}{n+1} \quad \text{for } j=0,\ldots,k. \end{aligned}$$

Then, we have

$$\frac{b_{k-1}-b_k}{a_{k-1}-a_k} = -\frac{1}{\lambda(n+1-k)} \le \frac{(k+1)\lambda-1}{(n-k)+\lambda(2k-n)} = \frac{b_k-b_{k+1}}{a_k-a_{k+1}}. \tag{22}$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

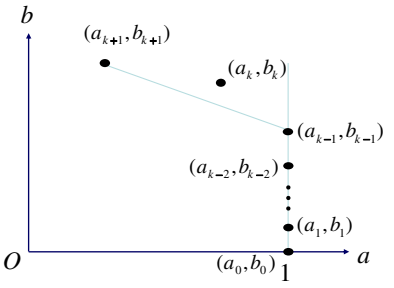


FIGURE 1. Positional relation among the points $(a_0, b_0), (a_1, b_1), \ldots, (a_k, b_k), (a_{k+1}, b_{k+1})$. Due to (22), the point (a_k, b_k) is located above the line segment between the points (a_{k-1}, b_{k-1}) and (a_{k+1}, b_{k+1}). This positional relation shows that the maximum in (23) is achieved by a probabilistic mixture among the three points $(a_{k-1}, b_{k-1}), (a_k, b_k), (a_{k+1}, b_{k+1})$.

Hence, the positional relation among the points $(a_0, b_0), (a_1, b_1), \ldots, (a_{k+1}, b_{k+1})$ are illustrated as Figure 1. The upper confidence limit $\overline{\epsilon}(\tilde{T}_k(\lambda), n, \delta)$ can be written as

$$\begin{aligned}\overline{\epsilon}(\tilde{T}_k(\lambda), n, \delta) &= \max_{(p_j)_{j=0}^{n+1}} \left\{ \frac{Q[\vec{p}](Y_{n+1}=1, \text{Pass } \tilde{T}_j(\lambda))}{Q[\vec{p}](\text{ Pass } \tilde{T}_j(\lambda))} \,\middle|\, Q[\vec{p}](\text{ Pass } \tilde{T}_j(\lambda)) \ge \delta \right\} \\ &= \max_{(p_j)_{j=0}^{n+1} \in \mathcal{P}_{k,n}} \left\{ \frac{Q[\vec{p}](Y_{n+1}=1, \text{Pass } \tilde{T}_j(\lambda))}{Q[\vec{p}](\text{ Pass } \tilde{T}_j(\lambda))} \,\middle|\, Q[\vec{p}](\text{ Pass } \tilde{T}_j(\lambda)) \ge \delta \right\}, \end{aligned} \tag{23}$$

where $\mathcal{P}_{k,n}$ is the set of distributions on $\{0, 1, \ldots, n, n+1\}$ such that $p_0 = p_1 = \cdots = p_{k-2} = p_{k+2} = \cdots = p_n = p_{n+1} = 0$.

When $a_{k+1} > \delta$, the maximum in (23) is realized with $p_{k-1} = p_k = 0$ and $p_{k+1} = 1$, and the maximum is 1. When $a_k > \delta \ge a_{k+1}$, the maximum is realized with $p_{k-1} = 0$, $p_k = (\delta(n+1) - (1-\lambda)(k+1))/(k + (1-\lambda)(n-2k))$ and $p_{k+1} = 1 - (\delta(n+1) - (1-\lambda)(k+1))/((k + (1-\lambda)(n-2k))$, and the maximum is $(\delta(n+1)(k - (1-\lambda)(k+1)) + (1-\lambda)^2(k+1)(n+1-k))/(\delta(n+1)(k + (1-\lambda)(n-2k)))$. When $1 = a_{k-1} \ge \delta \ge a_k$, the maximum is realized with $p_{k-1} = 1 - (1-\delta)(n+1-k)/(\lambda(n+1))$, $p_k = (1-\delta)(n+1-k)/[\lambda(n+1)]$, and $p_{k+1} = 0$, i.e. the maximum is $((k-1)\lambda(n+1-k) + (n+1)(1-\delta))/[\delta(n+1)\lambda(n+1-k)]$. Thus, we obtain (2).

4.3. Proof of (4)

In preparation for proving (4), we define the new quantity $\overline{\epsilon}^{\mathrm{c}}(T_0(\lambda), n, \delta) := 1 - \overline{\epsilon}(T_0(\lambda), n, \delta)$. In the following, we shall show (4) by proving that

$$\overline{\epsilon}^{\mathrm{c}}(T_0(\lambda), n, \delta_z) = \frac{(n-z+1)\lambda}{z + (n-z+1)\lambda} \quad \text{for all } z \in \{0, 1, \ldots, n+1\}. \tag{24}$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

Let $Q_z \in \mathcal{Q}_{n+1}$ be the permutation invariant distribution of variables $Y_1, \dots, Y_{n+1} \in \{0, 1\}$ that satisfy $\sum_{i=1}^{n+1} Y_i = z$. Then by a simple calculation we have

$$Q_z(L=0) = \begin{cases} \lambda^n & \text{when } z = n+1, \\ \frac{(n+1-z)\lambda^z + z\lambda^{z-1}}{n+1} = \delta_z & \text{when } 1 \le z \le n, \\ 1 & \text{when } z = 0, \end{cases}$$

$$Q_z(L=0, Y_{n+1}=0) = \begin{cases} 0 & \text{when } z = n+1, \\ \frac{(n+1-z)\lambda^z}{n+1} & \text{when } 1 \le z \le n, \\ 1 & \text{when } z = 0, \end{cases}$$

and, thus,

$$\frac{Q_z(L=0, Y_{n+1}=0)}{Q_z(L=0)} = \left\{ \begin{array}{ll} 0 & z = n+1 \\ \dfrac{(n+1-z)\lambda}{(n+1-z)\lambda + z} & 1 \le z \le n \\ 1 & z = 0 \end{array} \right\} = \frac{(n-z+1)\lambda}{z + (n-z+1)\lambda}. \tag{25}$$

In addition, we have

$$\overline{\epsilon}(T_0(\lambda), n, \delta) = \min_{\{p_j\}_{j=0}^{n+1}} \left\{ \frac{\sum_{j=0}^{n+1} p_j Q_j(L=0, Y_{n+1}=0)}{\sum_{j=0}^{n+1} p_j Q_j(L=0)} \,\middle|\, \sum_{j=0}^{n+1} p_j Q_j(L=0) \ge \delta \right\},$$

where p_j form a probability distribution on $[n+1]$.

Note that $Q_z(L=0)$, $Q_z(L=0, Y_{n+1}=0)$, and $Q_z(L=0, Y_{n+1}=0)/Q_z(L=0)$ all strictly decrease with z for $z \in \{0, 1, \dots, n+1\}$. Therefore, when $\delta = \delta_z$, the above minimum is realized with

$$p_j = \begin{cases} 1 \text{ when } j = z, \\ 0 \text{ otherwise.} \end{cases}$$

This fact together with (25) confirms (24) and completes the proof of (4).

5. Asymptotic Evaluations in the Linear Regime

We consider two types of asymptotic regimes. In the first type, the number l of allowed failures is proportional to n, and we take the limit $n \to \infty$. That is, we set the number of allowed failures l to be $\lceil svn \rceil$, where $0 \le s < 1$. In the i.i.d. setting, this regime leads to the central limit theorem. We call it the linear regime, and assume it in this section.

5.1. Preparations

To discuss the linear regime, we prepare the function $\psi(\delta)$ as

$$\psi(\delta) := \frac{\phi(\Phi^{-1}(\delta))}{\delta \Phi^{-1}(\delta)},$$

which takes a central role in our analysis on $\overline{\epsilon}(T_{\lceil svn \rceil}(\lambda), n, \delta)$. This quantity can be approximated as

$$\psi(\delta) \cong -B\left(1 - \mathrm{e}^{c\Phi^{-1}(\delta)}\right)^{-1},$$

where $B = 1.135$ and $c = 1.98/\sqrt{2}$ [26, 44].

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

When $x \geq 0$, the function $\eta(x) := x\Phi(x)/\phi(x)$ satisfies the condition $\mathrm{d}\eta(x)/\mathrm{d}x > 0$. Hence, $\psi(\delta)$ is strictly decreasing in $\delta > \frac{1}{2}$. Since $\lim_{\delta \to (\frac{1}{2})^+} \psi(\delta) = +\infty$ and $\lim_{\delta \to 1^-} \psi(\delta) = 0^+$, we can define ψ^{-1} on $\mathbb{R}^+$ with the range $(\frac{1}{2}, 1)$. Then, we define the function

$$\begin{aligned} C(\lambda, s, \delta) &:= \frac{\sqrt{(1-\lambda)s}(1-s)}{\sqrt{\lambda}} \Phi^{-1}(\delta) \left[\psi(\delta) - \frac{\lambda}{(1-\lambda)(1-s)} \right] \\ &= \sqrt{s}(1-s)\Phi^{-1}(\delta) \left[\sqrt{\frac{1-\lambda}{\lambda}} \psi(\delta) - \sqrt{\frac{\lambda}{1-\lambda}} \frac{1}{1-s} \right], \end{aligned} \tag{26}$$

where $\Phi(x)$ and $\phi(x)$ are defined in (10). When $s \leq 1$, the value $C(\lambda, s, \delta)$ is non-negative for $\delta \leq \psi^{-1}(\lambda/[(1-\lambda)(1-s)])$. Its minimization over λ is given as

$$\underline{C}(s, \delta) := \inf_{\lambda \in (0,1)} C(\lambda, s, \delta) = \begin{cases} 2\sqrt{\frac{-s(1-s)\Phi^{-1}(\delta)\phi(\Phi^{-1}(\delta))}{\delta}} & \text{when } \delta \leq \frac{1}{2}, \\ -\infty & \text{otherwise.} \end{cases} \tag{27}$$

When $\delta \leq \frac{1}{2}$, due to the relation of the arithmetic–geometric means, the above minimum value is attained when $\sqrt{(1-\lambda)/\lambda}\psi(\delta) = -\sqrt{\lambda/(1-\lambda)}(1/(1-s))$, i.e. $\lambda = (1-s)\psi(\delta)/((1-s)\psi(\delta) - 1)$ because the minimum is calculated as

$$\begin{aligned} &\min_{\lambda \in (0,1)} \sqrt{s}(1-s)\Phi^{-1}(\delta) \left[\sqrt{\frac{1-\lambda}{\lambda}} \psi(\delta) - \sqrt{\frac{\lambda}{1-\lambda}} \frac{1}{1-s} \right] \\ &= -2\sqrt{s}(1-s)\Phi^{-1}(\delta) \sqrt{\left[-\sqrt{\frac{1-\lambda}{\lambda}} \psi(\delta) \cdot \sqrt{\frac{\lambda}{1-\lambda}} \frac{1}{1-s} \right]} \\ &= -2\sqrt{s(1-s)}\Phi^{-1}(\delta)\sqrt{-\psi(\delta)} = 2\sqrt{\frac{-s(1-s)\Phi^{-1}(\delta)\phi(\Phi^{-1}(\delta))}{\delta}}. \end{aligned}$$

When $\delta > \frac{1}{2}$, the minimum is realized when $\lambda \to 1$.

5.2. Upper confidence limit

In the linear regime, according to (17) of Lemma1, when $0 < \delta \leq \frac{1}{2}$ and $0 \leq s < 1$, we have the following relation:

$$\overline{\epsilon}(T_{\lceil svn \rceil}(\lambda), n, \delta) > s.$$

Hence, our interest is the difference between $\overline{\epsilon}(T_{\lceil svn \rceil}(\lambda), n, \delta)$ and s.

Theorem 1. *We consider the linear regime.*

(a) *We fix $0 \leq s < 1$, $0 < \delta < 1$, and $0 < \lambda < 1$. When $s \leq \delta$ and n goes to infinity, we have*

$$\overline{\epsilon}(T_{\lceil sn \rceil}(0), n, \delta) = \frac{s}{\delta} + \left(\frac{1 - s + s^2 - \delta}{\delta(1-s)} + \frac{\upsilon_n}{\delta} \right) \frac{1}{n} + O\left(\frac{1}{n^2} \right), \tag{28}$$

where $\upsilon_n := \lceil sn \rceil - sn \leq 1$. When $s > \delta$ and n is sufficiently large, we have

$$\overline{\epsilon}(T_{\lceil sn \rceil}(0), n, \delta) = 1. \tag{29}$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

When n goes to infinity, we have

$$\overline{\epsilon}(T_{\lceil s\nu n\rceil}(\lambda), n, \delta) = s + C(\lambda, s, \delta)\sqrt{\frac{1}{n}} + O\left(\frac{1}{n}\right). \tag{30}$$

The coefficient $C(\lambda, s, \delta)$ *takes a non-negative value when* $\delta \le \psi^{-1}(\lambda/[\nu(1-s)])$. *That is, it is a positive number in a realistic situation.*

(b) *Suppose* $\delta = \mathrm{e}^{-rn}$ *for a certain constant* $r > 0$. *We fix* $0 \le s < 1$, $r > 0$, *and* $0 < \lambda < 1$. *When n goes to infinity we have*

$$\overline{\epsilon}(T_{\lceil sn\rceil}(0), n, \delta = \mathrm{e}^{-rn}) = 1, \tag{31}$$

$$\overline{\epsilon}(T_{\lceil s\nu n\rceil}(\lambda), n, \delta = \mathrm{e}^{-rn}) = \begin{cases} 1 & r > D(s\nu\|\nu), \\ E_{\lambda,s}(r) + o(1) & 0 < r \le D(s\nu\|\nu), \end{cases} \tag{32}$$

where

$$E_{\lambda,s}(r) := \frac{rt_{\mathrm{D}}(\nu s/r, \nu) - \nu s}{\lambda - \nu s + \nu r t_{\mathrm{D}}(\nu s/r, \nu)}.$$

The comparison between (30) and (28) leads to an advantage of our randomized test. Equation (30) shows that the upper confidence limit attains s for $\delta > 0$ and $\lambda > 0$ while (28) shows that it is impossible for $\delta > 0$ and $\lambda = 0$. That is, to achieve the upper confidence limit s, we need to choose non-zero λ.

Equations (29) and (31) are shown by the substitution of sn into k at (1) with the asymptotic expansion for n. The derivation of (28) is slightly complicated. Let $s_n = \lceil sn\rceil/n$; then $0 \le s_n - s = \upsilon_n/n \le 1/n$. The substitution of sn into k at (1) with the asymptotic expansion for n implies that

$$\overline{\epsilon}(T_{\lceil sn\rceil}(0), n, \delta) = \frac{s_n}{\delta} + \frac{1 - s_n + s_n^2 - \delta}{\delta(1 - s_n)}\frac{1}{n} + O\left(\frac{1}{n^2}\right), \tag{33}$$

Substituting $s_n = s + \upsilon_n/n$ in (33), we obtain (28). Equations (30) and (32) are shown in Section 9 after several preparations.

Equation (30) shows that the term $C(\lambda, s, \delta)n^{-1/2}$ is an extra value to guarantee the significance level δ. Hence, a smaller coefficient $C(\lambda, s, \delta)$ is better. That is, the minimizer

$$\lambda_{s,\delta} := \operatorname*{argmin}_{\lambda} C(\lambda, s, \delta) = \frac{(1-s)\psi(\delta)}{(1-s)\psi(\delta) - 1}$$

gives the asymptotically minimum upper confidence limit when s and $\delta \in (0, \frac{1}{2})$ are fixed. To see this characterization, we consider the minimum upper confidence limit:

$$\overline{\epsilon}_{\min}(s, n, \delta) := \min_{\lambda\in[0,1)} \overline{\epsilon}(T_{\lceil s\nu n\rceil}(\lambda), n, \delta).$$

Lemma 1 implies that

$$\overline{\epsilon}_{\min}(s, n, \delta) \ge \frac{\lceil \nu sn\rceil}{\nu n} \ge s \quad \text{for all } 0 < \delta \le \frac{1}{2}. \tag{34}$$

Equation (30) yields the following corollary for the minimum upper confidence limit.

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

Corollary 1. *We fix* $0 \le s < 1$, $0 < \delta < \frac{1}{2}$, *and* $0 < \lambda < 1$. *When n goes to infinity, we have*

$$\overline{\epsilon}_{\min}(s, n, \delta) = s + \underline{C}(s, \delta)\sqrt{\frac{1}{n}} + O\left(\frac{1}{n}\right). \tag{35}$$

Due to (31), when the significance level δ is exponentially small and $\lambda = 0$, the upper confidence limit takes the trivial value 1. That is, to achieve an exponentially small significance level δ, we need to choose non-zero λ, which is an advantage of our randomized test. Indeed, when the significance level δ is exponentially small value e^{-nr} and $\lambda > 0$, (32) shows the upper confidence limit. Hence, when s and r are fixed, the asymptotically minimum upper confidence limit equals the minimum $\min_\lambda \min\{1, E_{\lambda,s}(r)\}$, because the condition $E_{\lambda,s}(r) \ge 1$ is equivalent to the condition $r \ge D(s\nu\|\nu)$. Since we have $D(s\nu\|\nu) \to \infty$ when $\lambda \to 0^+$, this value is a non-trivial value.

In fact, when r approaches 0^+, $rt_{\mathrm{D}}(\nu s.r, \nu)$ approaches s^+, and $E_{\lambda,s}(r)$ approaches s^+. In addition, when r approaches $+\infty$, we have $E_{\lambda,s}(r) \ge 1$. This fact shows that the quantity $\overline{\epsilon}(T_{\lceil \nu sn \rceil}(\lambda), n, \delta = e^{-rn})$ takes values in $(s,1]$.

In contrast, we have the following proposition under the i.i.d. case. Under this case with the linear regime, we set k to be $\lceil sn \rceil$ with $0 \le s < 1$. According to Lemma 1, when $0 < \delta \le \frac{1}{2}$ we have

$$\overline{\epsilon}_{\text{i.i.d.}}(T_{\lceil sn \rceil}, n, \delta) > s. \tag{36}$$

Proposition 2. *We consider the linear regime.*

(a) *We fix* $0 \le s < 1$ *and* $0 < \delta < 1$. *When n goes to infinity, we have*

$$\overline{\epsilon}_{\text{i.i.d.}}(T_{\lceil sn \rceil}, n, \delta) = s - \Phi^{-1}(\delta)\sqrt{s(1-s)}\sqrt{\frac{1}{n}} + O\left(\frac{1}{n}\right). \tag{37}$$

(b) *Suppose* $\delta = e^{-rn}$ *for a certain constant* $r > 0$. *We fix* $0 \le s < 1$ *and* $r > 0$. *When n goes to infinity, we have*

$$\overline{\epsilon}_{\text{i.i.d.}}(T_{\lceil sn \rceil}, n, \delta = e^{-rn}) = \epsilon_D(s, r) + o(1). \tag{38}$$

Cases (a) and (b) immediately follow from the central limit theorem and large deviation principle, respectively. For the reader's convenience, Appendix E gives their proofs. A numerical comparison among (30), (35), and (37) is presented in Figure 2(b).

5.3. Minimum number of required observations *n*

Next, under the linear regime, we consider the minimum number of observations n required to guarantee the upper confidence limit ϵ at significance level δ. To this end, for $0 \le \lambda < 1$, $0 < \epsilon, \delta < 1$, and $0 \le s < 1$, we define the following values,

$$\begin{aligned} N_{\lambda,\mathrm{l}}(s, \epsilon, \delta) &:= \min\{n \ge \lceil \nu sn \rceil + 1 \mid \overline{\epsilon}(T_{\lceil s\nu n \rceil}(\lambda), n, \delta) \le \epsilon\}, \\ \underline{N}_{\mathrm{l}}(s, \epsilon, \delta) &:= \min_{\lambda \in [0,1)} N_{\lambda,\mathrm{l}}(s, \epsilon, \delta), \\ N_{\text{i.i.d.},\mathrm{l}}(s, \epsilon, \delta) &:= \min\{n \ge \lceil sn \rceil + 1 \mid \overline{\epsilon}_{\text{i.i.d.}}(T_{\lceil sn \rceil}, n, \delta) \le \epsilon\}. \end{aligned} \tag{39}$$

In the above notation, the subscript l expresses the linear regime. To make a fair comparison of the required number n, we need to choose the number of allowed failures l or k depending on λ and s in the above way.

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

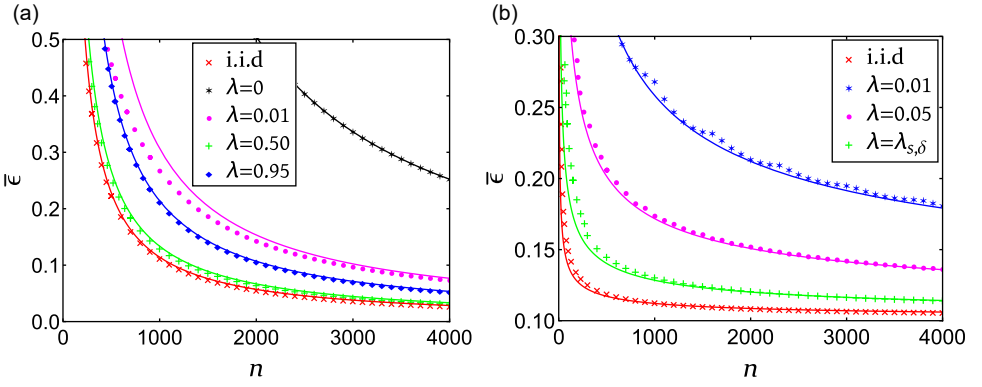


FIGURE 2. (a) Variations of upper confidence limits in the constant regime. Here $k_0 = 100$ and $\delta = 0.1$. The symbols '$\times$' represent $\overline{\epsilon}_{\text{i.i.d.}}(T_{k_0}, n, \delta)$, '$*$' represent $\overline{\epsilon}(T_{k_0}(0), n, \delta)$, whereas other symbols represent $\overline{\epsilon}(T_{\lceil \nu k_0 \rceil}(\lambda), n, \delta)$. The five curves from left to right represent the approximate formula in (66), (63), (63), (63), and (62), respectively (the terms $O(n^{-1})$ are ignored). (b) Variations of upper confidence limits in the linear regime. Here $s = 0.1$ and $\delta = 0.1$. The symbols '$\times$' represent $\overline{\epsilon}_{\text{i.i.d.}}(T_{\lceil sn \rceil}, n, \delta)$, and other symbols represent $\overline{\epsilon}(T_{\lceil svn \rceil}(\lambda), n, \delta)$. The four curves from left to right represent the approximate formula in (37), (35), (30), and (30), respectively (the terms $O(n^{-1})$ are ignored); $\overline{\epsilon}(T_{\lceil sn \rceil}(0), n, \delta)$ is not shown because its value is too large.

For the asymptotic evaluation of the above values, we have the following corollary.

Corollary 2. *Suppose* $0 < \lambda, \delta < 1$ *and* $0 \le s < \epsilon < 1$.

(a) *We fix* s, λ, *and* δ. *When* $0 < \delta \le \frac{1}{2}$ *and* ϵ *goes to* s^+, *we have*

$$N_{\lambda,1}(s, \epsilon, \delta) = \left(\frac{C(\lambda, s, \delta)}{\epsilon - s}\right)^2 + O\left(\frac{1}{\epsilon - s}\right), \tag{40}$$

$$\underline{N}_1(s, \epsilon, \delta) = \left(\frac{\underline{C}(s, \delta)}{\epsilon - s}\right)^2 + O\left(\frac{1}{\epsilon - s}\right), \tag{41}$$

$$N_{\text{i.i.d.},1}(s, \epsilon, \delta) = \frac{[\Phi^{-1}(\delta)]^2 s(1-s)}{(\epsilon - s)^2} + O\left(\frac{1}{\epsilon - s}\right), \tag{42}$$

where $C(\lambda, s, \delta)$ *and* $\underline{C}(s, \delta)$ *are defined in (26) and (27), respectively. We fix* s, δ *with the condition* $s < \delta \le \frac{1}{2}$. *We have*

$$\limsup_{\epsilon \to (\frac{s}{\delta})^+} \left| \left(\epsilon - \frac{s}{\delta}\right) N_{0,1}(s, \epsilon, \delta) - \frac{1 - s + s^2 - \delta}{\delta(1-s)} - \frac{1}{2\delta} \right| < \frac{1}{2\delta}. \tag{43}$$

(b) *When* ϵ, s, *and* λ *are fixed and* δ *goes to zero, we have*

$$N_{\lambda,1}(s, \epsilon, \delta) = \frac{1}{E^{-1}_{\lambda,s}(\epsilon)} \ln \delta^{-1} + o(\ln \delta^{-1}), \tag{44}$$

$$N_{\text{i.i.d.},1}(s, \epsilon, \delta) = \frac{1}{D(s\|\epsilon)} \ln \delta^{-1} + o(\ln \delta^{-1}). \tag{45}$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

Proof. The condition $0 < \delta \le \frac{1}{2}$ guarantees $C(\lambda, s, \delta) \ge 0$ and $\overline{\epsilon}(T_{\lceil svn \rceil}(\lambda), n, \delta) > s$. The definition in (39) guarantees that $N_{\lambda,1}(s, \epsilon, \delta)$ is non-increasing in ϵ. Due to (30) and $C(\lambda, s, \delta) \ge 0$, the equation $\epsilon = \overline{\epsilon}(T_{\lceil svn \rceil}(\lambda), n, \delta)$ guarantees $C(\lambda, s, \delta)\sqrt{\frac{1}{n}} = \epsilon - s + O(n^{-1})$, which implies that $n = C(\lambda, s, \delta)^2/(\epsilon - s)^2 + O\big((\epsilon - s)^{-1}\big)$. Hence, there exists an integer $n_0 > 0$ and some constant $C_0 > 0$ such that

$$\left| n - \frac{C(\lambda, s, \delta)^2}{(\epsilon - s)^2} \right| \le \frac{C_0}{\epsilon - s} \quad \text{for all } n \ge n_0.$$

On the other hand, due to Lemma 1, there exits an $s < \epsilon_0 < 1$ such that

$$\overline{\epsilon}(T_{\lceil svn \rceil}(\lambda), n, \delta) \ge \epsilon_0 \quad \text{for all } n \le n_0.$$

Therefore, for $\epsilon \in (s, \epsilon_0)$, we have

$$N_{\lambda,1}(s, \epsilon, \delta) = \frac{C(\lambda, s, \delta)^2}{(\epsilon - s)^2} + O\left(\frac{1}{\epsilon - s}\right).$$

Hence, we obtain (40). In the same way, we can show (41) from (34) and (35), and show (42) from (36) and (37). In addition, (43) follows from (18) and (28) of Lemma 1 in the same way because of $1 - s + s^2 \ge \frac{1}{2}$. By solving the equation $E_{\lambda,s}(r) = \epsilon$ for r, we obtain (44) from (32). Similarly, by solving $\epsilon_D(s, r) = \epsilon$ for r, we obtain (45) from (38). □

5.4. Maximum number of allowed failures *l*

In addition, we may consider the following quantities:

$$\begin{aligned} l_\lambda(n, \epsilon, \delta) &:= \max\{l \ge 0 \mid \overline{\epsilon}(T_l(\lambda), n, \delta) \le \epsilon\}, \\ l_{\text{i.i.d.}}(n, \epsilon, \delta) &:= \max\{l \ge 0 \mid \overline{\epsilon}_{\text{i.i.d.}}(T_l, n, \delta) \le \epsilon\}, \end{aligned} \tag{46}$$

which denote the maximum numbers of failures l allowed to guarantee the upper confidence limit ϵ at significance level δ.

Proposition 3. *Suppose* $0 < \lambda, \delta, \epsilon < 1$.

(a) *When* ϵ, λ, δ *are fixed and n goes to infinity, we have*

$$l_0(n, \epsilon, \delta) = \left\lfloor \delta\epsilon n - \frac{1 - \delta - \delta\epsilon + \delta^2\epsilon^2}{1 - \delta\epsilon} + O\left(\frac{1}{n}\right) \right\rfloor, \tag{47}$$

When $\delta \le \psi^{-1}\left(\frac{\lambda}{(1-\lambda)(1-s)}\right)$ *and n goes to infinity, we have*

$$l_\lambda(n, \epsilon, \delta) = \epsilon\nu n - C(\lambda, \epsilon, \delta)\nu\sqrt{n} + O(1), \tag{48}$$

When $\delta \le \frac{1}{2}$ *and n goes to infinity, we have*

$$l_{\text{i.i.d.}}(n, \epsilon, \delta) = \epsilon n + \Phi^{-1}(\delta)\sqrt{\epsilon(1 - \epsilon)n} + O(1), \tag{49}$$

where $C(\lambda, s, \delta)$ *is defined in (26).*

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

(b) *Suppose $\delta = \mathrm{e}^{-rn}$ for a certain constant $r > 0$. When r is fixed and n goes to infinity, we have*

$$l_\lambda(n, \epsilon, \mathrm{e}^{-rn}) = \nu \bar{E}^{-1}_{\lambda,r}(\epsilon)n + o(n) \quad \text{for } r \le \frac{-\lambda\epsilon \ln \lambda}{1 - \nu\epsilon}, \tag{50}$$

$$l_{\text{i.i.d.}}(n, \epsilon, \mathrm{e}^{-rn}) = s_D(\epsilon, r)n + o(n) \quad \text{for } r \le -\ln(1 - \epsilon), \tag{51}$$

where $s_D(\epsilon, r)$ is defined in (13), $\bar{E}_{\lambda,r}(s) := E_{\lambda,s}(r)$, and $\bar{E}^{-1}_{\lambda,r}(\epsilon)$ denotes the inverse function of $\bar{E}_{\lambda,r}$.

Proof. First, we prove (47). The condition $\overline{\epsilon}(T_l, n, \delta) \le \epsilon$ is equivalent to the conditions

$$l^2 - [(1 + \epsilon\delta)n + \epsilon\delta]l - (n + 1)(1 - \delta - \epsilon\delta n) \ge 0 \tag{52}$$

and $l \le \delta(n + 1) - 1$. The condition in (52) is equivalent to

$$l \le \delta\epsilon n - \frac{1 - \delta - \delta\epsilon + \delta^2\epsilon^2}{1 - \delta\epsilon} + O\left(\frac{1}{n}\right) \quad \text{or} \quad l \ge n + \frac{1 - \delta}{1 - \delta\epsilon} + O\left(\frac{1}{n}\right). \tag{53}$$

However, the second case in (53) is not allowed due to the condition $l \le \delta(n + 1) - 1$. Hence, we have only the first case in (53). Since $l_0(n, \epsilon, \delta)$ is the maximum integer that satisfies the first inequality in (53), we have (47).

Next, we prove (48). When l is proportional to n, (30) implies that

$$\overline{\epsilon}(T_l(\lambda), n, \delta) = \frac{l}{\nu n} + C\left(\lambda, \frac{l}{\nu n}, \delta\right)\sqrt{\frac{1}{n}} + O\left(\frac{1}{n}\right). \tag{54}$$

To determine $l_\lambda(n, \epsilon, \delta)$, we set the right-hand side of (54) to be ϵ, that is,

$$\frac{l}{\nu n} + C\left(\lambda, \frac{l}{\nu n}, \delta\right)\sqrt{\frac{1}{n}} + O\left(\frac{1}{n}\right) = \epsilon. \tag{55}$$

This implies that $l = \nu n\epsilon + x$, where $x = o(n)$ is some small quantity. By plugging $l = \nu n\epsilon + x$ into (55), we get

$$\frac{x}{\nu} + C(\lambda, \epsilon + o(1), \delta)\sqrt{n} + O(1) = 0.$$

Hence, we have $x = -C(\lambda, \epsilon, \delta)\nu\sqrt{n} + o(\sqrt{n})$ and $l = \nu n\epsilon - C(\lambda, \epsilon, \delta)\nu\sqrt{n} + y$, where $y = o(\sqrt{n})$. By plugging this relation into (55) again, we can deduce that $y = O(1)$. This confirms (48). Equation (49) can be derived in a similar way.

Finally, we prove (50). Solving the equation $\bar{E}_{\lambda,r}(s) = E_{\lambda,s}(r) = \epsilon$ for s, we obtain $l/[(1 - \lambda)n] = \bar{E}^{-1}_{\lambda,r}(\epsilon)$ from (32). Note that $E_{\lambda,s}(r)$ is increasing in s and this equation has a solution for $s \ge 0$ when the constraint $E_{\lambda,0}(r) \le \epsilon$, i.e. $r \le -\lambda\epsilon \ln \lambda/(1 - \nu\epsilon)$ holds. Hence, we obtain (50) under the condition $r \le -\lambda\epsilon \ln \lambda/(1 - \nu\epsilon)$. In the same way, we obtain (51) from (38). □

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

6. Power and Detection Probability under the i.i.d. Distribution

Until now, to cover soundness, we have discussed how to design our test with the randomization parameter λ. By (46), if we choose the number of allowed failures $l = l_\lambda(n, \epsilon, \delta)$, then we can guarantee that the conditional probability $\mathrm{P}(Y_{n+1} = 1 | L \le l)$ is not greater than the upper confidence limit ϵ at significance level δ. However, we have not discussed how our test efficiently detects the event guaranteeing the above condition. In the following, to cover completeness, we consider the power, i.e. the detection probability $\mathrm{P}(L \le l)$ of our test under random sampling without replacement. Since completeness cares only the i.i.d. case, we assume that the true distribution belongs to the set $\mathcal{Q}_{\mathrm{i.i.d.},n+1}$ in this section. We assume that the number n of observations is sufficiently large for the normal approximation, and that the true distribution is given as $P^{n+1}_{\theta_0 + t/\sqrt{n}}$. For $\epsilon > \theta_0 + t/\sqrt{n}$, a higher power, i.e. a higher detection probability is preferred.

When $\epsilon = \theta_0 + e/\sqrt{n}$ and $0 < \lambda < 1$, the maximum number of allowed failures is

$$l_\lambda\left(n, \theta_0 + \frac{e}{\sqrt{n}}, \delta\right) = \theta_0(1-\lambda)n + e(1-\lambda)\sqrt{n} - C(\lambda, \theta_0, \delta)(1-\lambda)\sqrt{n} + O(1)$$

due to (48). To make a fair comparison, we focus on the threshold detecting the desired event with respect to a random variable asymptotically subject to the standard normal distribution under the normal approximation.

First, we fix $\lambda = 0$. Then, the maximum number of allowed failures is

$$l_0\left(n, \theta_0 + \frac{e}{\sqrt{n}}, \delta\right) = \delta\theta_0 n + \delta e\sqrt{n} + O(1).$$

The random variable $K^{(n)} := \left(K - \left(\theta_0 + t/\sqrt{n}\right) n\right)/\sqrt{n\theta_0(1-\theta_0)}$ is subject to the standard normal distribution under the distribution $P^{n+1}_{\theta_0 + t/\sqrt{n}}$. The threshold for the variable $K^{(n)}$ is given as

$$\frac{l_0\left(n, \theta_0 + \frac{e}{\sqrt{n}}, \delta\right) - \left(\theta_0 + \frac{t}{\sqrt{n}}\right) n}{\sqrt{n\theta_0(1-\theta_0)}} = -\frac{(1-\delta)\sqrt{\theta_0 n}}{\sqrt{1-\theta_0}} + \frac{\delta e - t}{\sqrt{\theta_0(1-\theta_0)}} + O\left(\frac{1}{\sqrt{n}}\right). \tag{56}$$

Since the probability Φ(56) goes to zero, the power, i.e. the detection probability converges to zero. That is, it is almost impossible to guarantee that the conditional probability $\mathrm{P}(Y_{n+1} = 1 | L \le l)$ is not greater than the upper confidence limit $\epsilon = \theta_0 + e/\sqrt{n}$ at significance level δ when $\lambda = 0$ and the true distribution is $P^{n+1}_{\theta_0 + t/\sqrt{n}}$, regardless of the values of e, t, and δ. This can be considered as a serious defect of random sampling without replacement.

However, our randomized test with non-zero $\lambda > 0$ resolves this problem as follows. In this case, the random variable $L^{(n)}(\lambda) := \left(L - \left(\theta_0 + t/\sqrt{n}\right)(1-\lambda)n\right)/\sqrt{n\theta_0(1-\lambda)(1-\theta_0(1-\lambda))}$ is subject to the standard normal distribution under the normal approximation when the true distribution is $P^{n+1}_{\theta_0 + t/\sqrt{n}}$. The threshold for the random variable $L^{(n)}(\lambda)$ is given as

$$\begin{aligned}
&\frac{l_\lambda\left(n, \theta_0 + \frac{e}{\sqrt{n}}, \delta\right) - \left(\theta_0 + \frac{t}{\sqrt{n}}\right)(1-\lambda)n}{\sqrt{n\theta_0(1-\lambda)(1-\theta_0(1-\lambda))}} \\
&= \frac{(e-t)(1-\lambda)}{\sqrt{\theta_0(1-\lambda)(1-\theta_0(1-\lambda))}} - \frac{C(\lambda, \theta_0, \delta)(1-\lambda)}{\sqrt{\theta_0(1-\lambda)(1-\theta_0(1-\lambda))}} + O\left(\frac{1}{\sqrt{n}}\right) \\
&= \frac{\sqrt{1-\lambda}}{\sqrt{\theta_0(1-\theta_0(1-\lambda))}}((e-t) - C(\lambda, \theta_0, \delta)) + O\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

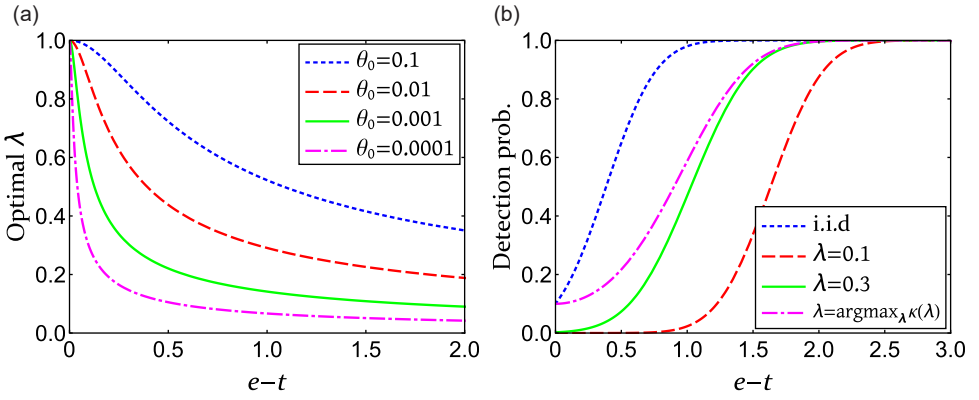


FIGURE 3. (a) Variations of $\mathrm{argmax}_{\lambda\in(0,1)}\,\kappa(\lambda)$ with $e-t$ for different θ_0. Here we choose $\delta=0.1$. The curves are calculated by solving (58) based on Lemma 2. (b) Variations of the power with $e-t$. Here $\theta_0=0.1$ and $\delta=0.1$. The first curve from the left expresses (59), which corresponds to the i.i.d. case; the other three curves express (57). The power $\Phi(56)$ is not shown because its value is too small.

That is, the limiting value of the power is

$$\Phi\left(\frac{\sqrt{1-\lambda}}{\sqrt{\theta_0(1-\theta_0(1-\lambda))}}\big((e-t)-C(\lambda,\theta_0,\delta)\big)\right), \tag{57}$$

which is a strictly positive value. In other words, with the above probability, our randomized test guarantees that the conditional probability $\mathrm{P}(Y_{n+1}=1|L\le l)$ is not greater than the upper confidence limit $\epsilon=\theta_0+e/\sqrt{n}$ at significance level δ when the true distribution is $P^{n+1}_{\theta_0+t/\sqrt{n}}$. This can be considered as a significant advantage of our randomized test.

In addition, to achieve the power p_0, the parameters e, t, δ, and λ need to satisfy the condition

$$\frac{\sqrt{1-\lambda}}{\sqrt{\theta_0(1-\theta_0(1-\lambda))}}\big((e-t)-C(\lambda,\theta_0,\delta)\big)\ge\Phi^{-1}(p_0)+O\left(\frac{1}{\sqrt{n}}\right),$$

which is equivalent to

$$e-t\ge C(\lambda,\theta_0,\delta)+\frac{\sqrt{\theta_0(1-\theta_0(1-\lambda))}}{\sqrt{1-\lambda}}\Phi^{-1}(p_0)+O\left(\frac{1}{\sqrt{n}}\right).$$

To achieve the power p_0, the difference $e-t$ needs to satisfy the above condition.

In the following, we maximize the power (57) by changing λ when the parameters θ_0, e, t, and δ are fixed. That is, we maximize $(e-t)\sqrt{1-\lambda}/\sqrt{\theta_0(1-\theta_0(1-\lambda))}-C(\lambda,\theta_0,\delta)\sqrt{1-\lambda}/\sqrt{\theta_0(1-\theta_0(1-\lambda))}$ with respect to λ. This term is written as $\kappa(\lambda):=(\alpha\sqrt{\lambda}+\beta(1-\lambda)/\sqrt{\lambda}+\gamma\sqrt{1-\lambda})/(1-\theta_0(1-\lambda))^{1/2}$ with $\alpha=\Phi^{-1}(\delta)$, $\beta=-\Phi^{-1}(\delta)\psi(\delta)(1-\theta_0)$, and $\gamma=(e-t)/\sqrt{\theta_0}$. Then, the optimal choice of λ is characterized by the following lemma (see Appendix F for a proof) whereas the optimal λ is numerically given in Figure 3(a).

Lemma 2. *For $\lambda\in(0,1)$, the equation $\kappa'(\lambda)=0$ is equivalent to the following polynomial equation with degree 3:*

$$(1-\lambda)[\Phi^{-1}(\delta)(1-\theta_0)]^2\Big[\,[1+(1+\theta_0)\psi(\delta)]\,\lambda+(1-\theta_0)\psi(\delta)\Big]^2=\frac{(e-t)^2}{\theta}\lambda^3. \tag{58}$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

For $\delta < \frac{1}{2}$, (58) has but only one solution in (0, 1). If we denote the solution by λ_0 then we have

$$\max_{\lambda\in(0,1)} \kappa(\lambda) = \kappa(\lambda_0).$$

In contrast, when we employ the above test under the i.i.d. assumption, the maximum number of allowed failures is

$$l_{\text{i.i.d.}}\left(n, \theta_0 + \frac{e}{\sqrt{n}}, \delta\right) = \theta_0 n + e\sqrt{n} + \Phi^{-1}(\delta)\sqrt{\theta_0(1-\theta_0)}\sqrt{n} + O(1).$$

The threshold for the variable $K^{(n)}$ is given as

$$\frac{l_{\text{i.i.d.}}\left(n, \theta_0 + \frac{e}{\sqrt{n}}, \delta\right) - \left(\theta_0 + \frac{t}{\sqrt{n}}\right) n}{\sqrt{n\theta_0(1-\theta_0)}} = \frac{e-t}{\sqrt{\theta_0(1-\theta_0)}} + \Phi^{-1}(\delta) + O\left(\frac{1}{\sqrt{n}}\right).$$

That is, the limiting value of the power is

$$\Phi\left(\frac{e-t}{\sqrt{\theta_0(1-\theta_0)}} + \Phi^{-1}(\delta)\right). \tag{59}$$

We compare the power in (57) and (59) as follows while their numerical comparison is given in Figure 3(b). Their magnitude relation follows from the magnitude relation of their input values. The first and second terms of their input values are evaluated as

$$\frac{e-t}{\sqrt{\theta_0(1-\theta_0)}} \geq \frac{\sqrt{1-\lambda}}{\sqrt{\theta_0(1-\theta_0(1-\lambda))}}(e-t), \tag{60}$$

$$\Phi^{-1}(\delta) \geq -\frac{\sqrt{1-\lambda}\, C(\lambda, \theta_0, \delta)}{\sqrt{\theta_0(1-\theta_0(1-\lambda))}} = \frac{\sqrt{\theta_0}(1-\theta_0)\Phi^{-1}(\delta)}{\sqrt{\theta_0(1-\theta_0(1-\lambda))}}\left[\frac{\sqrt{\lambda}}{1-\theta_0} - \frac{1-\lambda}{\sqrt{\lambda}}\psi(\delta)\right]. \tag{61}$$

Since Φ is monotonic increasing, their combination yields the relation (57) $\leq$ (59), i.e. the i.i.d. assumption realizes a larger power. This inequality cannot be saturated because the equality in (60) holds in the limit $\lambda \to 0$, but the equality in (61) holds in the limit $\lambda \to 1$.

7. Asymptotic Evaluations in the Constant Regime

In this section, we fix the number of allowed failures l to be a certain constant and take the limit $n \to \infty$. In the i.i.d. setting, this regime leads to the law of small numbers, i.e. Poisson limit theorem. We call it the constant regime. In this section, we assume the constant regime, i.e. the number of allowed failures l is fixed to be $\lceil \nu k_0 \rceil$ depending on $1 > \lambda \geq 0$ and an integer $k_0 \geq 0$ because the average of L is proportional to ν.

7.1. Upper confidence limit

For the asymptotic evaluation of $\overline{\epsilon}(T_l(\lambda), n, \delta)$, we define the coefficient

$$G(l, \delta, \lambda) := \frac{(B_{z_*,l} - \delta)z^* B_{z_*,l} + (\delta - B_{z^*,l})z_* B_{z_*-1,l}}{\delta \Delta_{z_*,l}},$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

where z^* and z_* are shorthand for $z^*(l, \delta, \lambda)$ and $z_*(l, \delta, \lambda)$, respectively, which are defined in (9). Then, we have the following theorem.

Theorem 2. *We consider the constant regime. We fix* $0 < \lambda < 1$ *and* $k_0 \in \mathbb{Z}^{\geq 0}$. *Then the following statements hold.*

(a) *When* $0 < \delta < 1$ *is fixed and n goes to infinity, the upper confidence limit has the following behavior*

$$\overline{\epsilon}(T_{k_0}(0), n, \delta) = \frac{k_0 + 1 - \delta}{\delta} n^{-1} + o(n^{-1}), \tag{62}$$

$$\overline{\epsilon}(T_{\lceil \nu k_0 \rceil}(\lambda), n, \delta) = G(\lceil \nu k_0 \rceil, \delta, \lambda) n^{-1} + o(n^{-1}). \tag{63}$$

(b) *Suppose* $\delta = \mathrm{e}^{-rn}$ *for a certain constant* $r > 0$. *When r is fixed and n goes to infinity, the upper confidence limit has the following behavior*

$$\overline{\epsilon}(T_{k_0}(0), n, \delta = \mathrm{e}^{-rn}) = 1, \tag{64}$$

$$\overline{\epsilon}(T_{\lceil \nu k_0 \rceil}(\lambda), n, \delta = \mathrm{e}^{-rn}) = \begin{cases} \frac{r}{r + \lambda(\ln \lambda^{-1} - r)} + o(1) & 0 < r \leq \ln \lambda^{-1}, \\ 1 & r > \ln \lambda^{-1}. \end{cases} \tag{65}$$

Equations (62) and (64) follow from (1) with the limit $n \to \infty$. Other equations will be shown in Appendix N after several preparations.

In contrast, we have the following proposition under the i.i.d. case.

Proposition 4. *We consider the constant regime. We fix* $k_0 \in \mathbb{Z}^{\geq 0}$.

(a) *When* $0 < \delta < 1$ *is fixed and n goes to infinity, we have*

$$\overline{\epsilon}_{\text{i.i.d.}}(T_{k_0}, n, \delta) = t_{\mathrm{P}}(k_0, \delta) n^{-1} + o(n^{-1}). \tag{66}$$

(b) *Suppose* $\delta = \mathrm{e}^{-rn}$ *for a certain constant* $r > 0$. *When r is fixed and n goes to infinity, we have*

$$\overline{\epsilon}_{\text{i.i.d.}}(T_{k_0}, n, \delta = \mathrm{e}^{-rn}) = 1 - \mathrm{e}^{-r} + o(1). \tag{67}$$

Cases (a) and (b) immediately follow from the law of small numbers and the large deviation principle, respectively. For the reader's convenience, we give their proofs in Appendix G. A numerical comparison among (62), (63), and (66) is presented as Figure 2(a).

7.2. Minimum number of required observations *n*

Next, we consider the minimum number of observations n required to guarantee the upper confidence limit ϵ at significance level δ. For $0 \leq \lambda < 1$, $0 < \epsilon, \delta < 1$, and $k, l \in \mathbb{Z}^{\geq 0}$, we define the following values:

$$N_{\lambda,\mathrm{c}}(l, \epsilon, \delta) := \min\{n \in \mathbb{Z}^{\geq l+1} \mid \overline{\epsilon}(T_l(\lambda), n, \delta) \leq \epsilon\},$$
$$N_{\text{i.i.d.},\mathrm{c}}(k, \epsilon, \delta) := \min\{n \in \mathbb{Z}^{\geq k+1} \mid \overline{\epsilon}_{\text{i.i.d.}}(T_k, n, \delta) \leq \epsilon\}.$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

In the above notation, the subscript c expresses the constant regime.

For the asymptotic evaluation of $N_{\lambda,\mathrm{c}}(l, \epsilon, \delta)$, we have the following result.

Corollary 3. *Suppose* $0 < \lambda, \epsilon, \delta < 1$ *and* $k_0 \in \mathbb{Z}^{\geq 0}$. *Then the following statements hold.*

(a) *When* k_0, δ, *and* λ *are fixed and* ϵ *goes to zero, we have*

$$N_{0,\mathrm{c}}(k_0, \epsilon, \delta) = \frac{k_0 + 1 - \delta}{\delta\epsilon} + o(\epsilon^{-1}), \tag{68}$$

$$N_{\lambda,\mathrm{c}}(\lceil \nu k_0 \rceil, \epsilon, \delta) = G(\lceil \nu k_0 \rceil, \delta, \lambda)\epsilon^{-1} + o(\epsilon^{-1}), \tag{69}$$

$$N_{\mathrm{i.i.d.},\mathrm{c}}(k_0, \epsilon, \delta) = t_{\mathrm{P}}(k_0, \delta)\epsilon^{-1} + o(\epsilon^{-1}). \tag{70}$$

(b) *When* k_0, ϵ, *and* λ *are fixed and* δ *goes to zero, we have*

$$N_{0,\mathrm{c}}(k_0, \epsilon, \delta) = \frac{k_0 + 1}{\delta\epsilon} + o(\delta^{-1}), \tag{71}$$

$$N_{\lambda,\mathrm{c}}(\lceil \nu k_0 \rceil, \epsilon, \delta) = \frac{1 - \nu\epsilon}{\lambda\epsilon \ln \lambda^{-1}} \ln \delta^{-1} + o(\ln \delta^{-1}), \tag{72}$$

$$N_{\mathrm{i.i.d.},\mathrm{c}}(k_0, \epsilon, \delta) = \frac{\ln \delta^{-1}}{\ln (1 - \epsilon)^{-1}} + o(\ln \delta^{-1}). \tag{73}$$

In case (b), choosing non-zero λ improves the required number $N_{\lambda,\mathrm{c}}(k_0, \epsilon, \delta)$ exponentially.

Proof. Equations (68), (69), and (70) follow from (62), (63), and (66), respectively. When δ^{-1} and n go to infinity, (71) follows from (1).

Equation (72) can be shown from (65) as follows. We consider the following equation for r:

$$\lim_{n\to\infty} \overline{\epsilon}(T_{\lceil \nu k_0 \rceil}(\lambda), n, \delta = \mathrm{e}^{-rn}) = \frac{r}{r + \lambda(\ln \lambda^{-1} - r)} = \epsilon.$$

The solution is $r = \lambda\epsilon \ln \lambda^{-1}/(1 - \nu\epsilon)$. Since $\delta = \mathrm{e}^{-nr}$, we obtain (72). In the same way, solving the equation $1 - \mathrm{e}^{-r} = \epsilon$ for r, we obtain (73) from (67). □

As shown in Appendix H after several preparations, we have the following lemma.

Lemma 3. *Suppose* $0 < \lambda, \epsilon, \delta < 1$ *and* $k_0 \in \mathbb{Z}^{\geq 0}$. *Then*

$$\lim_{\delta,\epsilon\to 0} \frac{\epsilon N_{\lambda,\mathrm{c}}(\lceil \nu k_0 \rceil, \epsilon, \delta)}{\ln \delta^{-1}} = \frac{1}{\lambda \ln \lambda^{-1}},$$

where the order of the two limits $\epsilon \to 0$ *and* $\delta \to 0$ *does not matter.*

The function $1/(\lambda \ln \lambda^{-1})$ takes the minimum value e when $\lambda = 1/\mathrm{e}$. Hence, the choice $\lambda = 1/\mathrm{e}$ is optimal under the constant regime.

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

8. Computation of the Upper Confidence Limit $\overline{\epsilon}(T_l(\lambda), n, \delta)$

8.1. Region approach

We extend the idea based on the region used in Section 4.2 to the case with the test $T_l(\lambda)$. In order to determine the upper confidence limit $\overline{\epsilon}(T_l(\lambda), n, \delta)$, we introduce the quantities $p_l(Q)$ and $f_l(Q)$ for $0 \le \lambda < 1$, $l \in \mathbb{Z}^{\ge 0}$, $n \in \mathbb{Z}^{\ge l+1}$, and $Q \in \mathcal{Q}_{n+1}$ as

$$p_l(Q) := Q(L \le l), \quad f_l(Q) := Q(Y_{n+1} = 0, L \le l).$$

Then, instead of $\overline{\epsilon}(T_l(\lambda), n, \delta)$, we discuss $\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta) := 1 - \overline{\epsilon}(T_l(\lambda), n, \delta)$. Define the two-dimensional region $\mathcal{R}_{l,n,\lambda}$ as

$$\mathcal{R}_{l,n,\lambda} := \left\{ \big(p_l(Q), f_l(Q)\big) \in \mathbb{R}_+^2 \;\middle|\; Q \in \mathcal{Q}_{n+1} \right\}. \tag{74}$$

This region $\mathcal{R}_{l,n,\lambda}$ is convex because $p_l(Q)$ and $f_l(Q)$ are both linear in Q, and the set $\mathcal{Q}_{n+1}$ is convex. The geometric picture of $\mathcal{R}_{l,n,\lambda}$ is helpful to understanding and calculating $\overline{\epsilon}(T_l(\lambda), n, \delta)$. In particular, $\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta)$ can be determined for all $0 < \delta \le 1$ once the lower boundary of $\mathcal{R}_{l,n,\lambda}$ is known, because it has the following form

$$\begin{aligned} \overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta) &= \min_{Q \in \mathcal{Q}_{n+1}} \{ Q(Y_{n+1} = 0 \mid L \le l) \mid Q(L \le l) \ge \delta \} \\ &= \min_{Q \in \mathcal{Q}_{n+1}} \left\{ \frac{f_l(Q)}{p_l(Q)} \;\middle|\; p_l(Q) \ge \delta \right\}. \end{aligned} \tag{75}$$

To characterize $\mathcal{R}_{l,n,\lambda}$, we define the random variable $Z := \sum_{i=1}^{n+1} Y_i$. Then any $Q \in \mathcal{Q}_{n+1}$ is completely determined by the $n+1$ probabilities $\{Q(Z = z)\}_{z=0}^{n}$, and the joint distribution $P_{Y_1,\ldots,Y_{n+1}}$ of $Y_1, \ldots, Y_{n+1}$ is given as

$$P_{Y_1,\ldots,Y_{n+1}}(y_1, \ldots, y_{n+1}) = \binom{n+1}{\sum_{i=1}^{n+1} y_i}^{-1} Q\left(Z = \sum_{i=1}^{n+1} y_i\right).$$

For integer $0 \le z \le n+1$, we define

$$h_z(l, n, \lambda) := Q(L \le l \mid Z = z) = \begin{cases} 1 & 0 \le z \le l, \\ \dfrac{(n - z + 1)B_{z,l} + zB_{z-1,l}}{n+1} & l+1 \le z \le n+1, \end{cases} \tag{76}$$

$$g_z(l, n, \lambda) := Q(Y_{n+1} = 0, L \le l \mid Z = z) = \begin{cases} \dfrac{n - z + 1}{n+1} & 0 \le z \le l, \\ \dfrac{(n - z + 1)B_{z,l}}{n+1} & l+1 \le z \le n+1, \end{cases} \tag{77}$$

where $B_{z,l}$ is shorthand for $B_{z,l}(\nu)$ as defined in (7), and $B_{z,l}(\nu) = 1$ for $l \ge z$. The second equality of (76) and the second equality of (77) are proved in Appendix 1. Then, we have

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

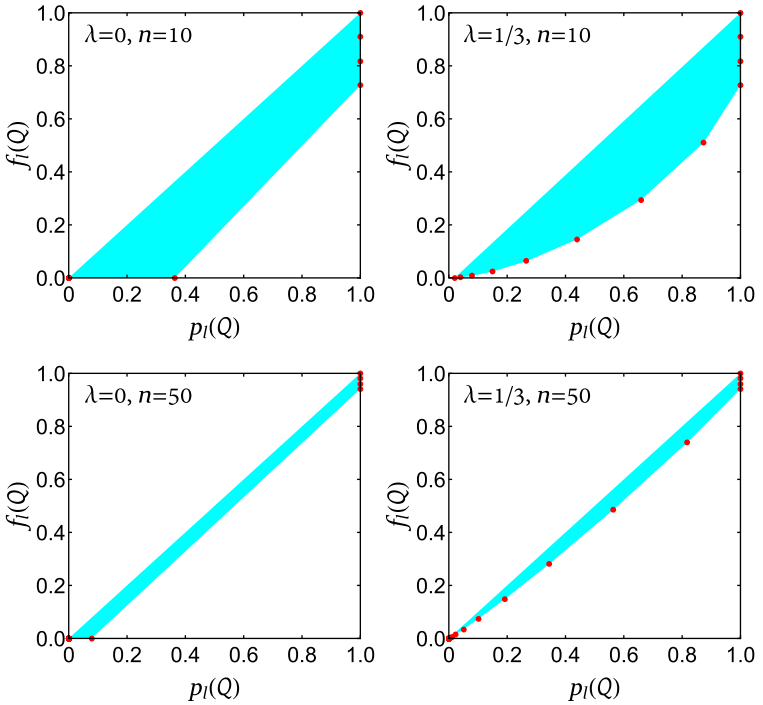


FIGURE 4. The region $\mathcal{R}_{l,n,\lambda}$ composed of points $(p_l(Q), f_l(Q))$ as defined in (74). The region is the convex hull of the points in (78), which are highlighted as dots. Here we choose $l = 3$ for all plots.

$$Q(L \le l) = \sum_{z=0}^{n+1} Q(Z = z) Q(L \le l \mid Z = z) = \sum_{z=0}^{n+1} Q(Z = z) h_z(l, n, \lambda),$$

$$Q(Y_{n+1} = 0, L \le l) = \sum_{z=0}^{n+1} Q(Z = z) Q(Y_{n+1} = 0, L \le l \mid Z = z)$$

$$= \sum_{z=0}^{n+1} Q(Z = z) g_z(l, n, \lambda).$$

Therefore, any element of the region $\mathcal{R}_{l,n,\lambda}$ can be written as a convex combination of the $n + 2$ points $\{(h_z(l, n, \lambda), g_z(l, n, \lambda))\}_{z=0}^{n+1}$. That is, we obtain the following proposition.

Proposition 5. *The two-dimensional region $\mathcal{R}_{l,n,\lambda}$ (see Figure 4) is the convex hull of the following set of points*

$$\left\{\left(h_z(l, n, \lambda), g_z(l, n, \lambda)\right)\right\}_{z=0}^{n+1}. \tag{78}$$

According to this proposition, $\mathcal{R}_{l,n,\lambda}$ is a polygon, as illustrated in Figure 4. Now, the original complicated problem of determining the region $\mathcal{R}_{l,n,\lambda}$ has been transformed into the intuitive task of studying the geometric relation between the $n + 2$ points given in (78).

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

8.2. Case with $\lambda > 0$

In this subsection we discuss $\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta)$ in the case $0 < \lambda < 1$.

As the first step, we clarify that the lower boundary of the region $\mathcal{R}_{l,n,\lambda}$ defined in (74) consists of $n - l + 1$ line segments, as shown in the following lemmas (see Appendix J for proofs).

Lemma 4. *Suppose* $0 < \lambda < 1$, $l, z \in \mathbb{Z}^{\geq 0}$, *and* $n \in \mathbb{Z}^{\geq l+1}$. *Then* $h_z(l, n, \lambda) > 0$; $g_z(l, n, \lambda) \geq 0$ *for all* $z \leq n + 1$; $g_z(l, n, \lambda) < 0$ *for all* $z \geq n + 2$. *In addition,* $h_z(l, n, \lambda)$ *strictly decreases with* z *for* $l \leq z \leq n + 1$; $g_z(l, n, \lambda)$ *strictly decreases with* z *for* $0 \leq z \leq n + 1$.

Lemma 5. *Suppose* $0 < \lambda < 1$, $l, z \in \mathbb{Z}^{\geq 0}$, *and* $n \in \mathbb{Z}^{\geq l+1}$. *Then for* $l \leq z \leq n + 1$, *the point* $\big(h_z(l, n, \lambda), g_z(l, n, \lambda)\big)$ *is an extremal point of* $\mathcal{R}_{l,n,\lambda}$ *and belongs to the lower boundary of* $\mathcal{R}_{l,n,\lambda}$.

Next, we shall determine the exact value of $\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta)$. By Lemma 4, we have $B_{n,l} = h_{n+1}(l, n, \lambda) \leq h_z(l, n, \lambda) \leq 1$ for $z \in \{0, \ldots, n + 1\}$. For $B_{n,l} < \delta \leq 1$, let $\hat{z}$ be the largest integer z such that $h_z(l, n, \lambda) \geq \delta$. For $z \in \{l, l + 1, \ldots, n\}$, define

$$\kappa_z(l, n, \delta, \lambda) := \frac{\delta - h_{z+1}(l, n, \lambda)}{h_z(l, n, \lambda) - h_{z+1}(l, n, \lambda)}, \tag{79}$$

$$\tilde{\zeta}_\lambda(l, n, \delta, z) := [1 - \kappa_z(l, n, \delta, \lambda)]g_{z+1}(l, n, \lambda) + \kappa_z(l, n, \delta, \lambda)g_z(l, n, \lambda). \tag{80}$$

The following theorem provides the analytical expression of $\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta)$.

Theorem 3. *Suppose* $0 < \lambda < 1$, $0 < \delta \leq 1$, $l \in \mathbb{Z}^{\geq 0}$, *and* $n \in \mathbb{Z}^{\geq l+1}$. *Then we have*

$$\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta) = \begin{cases} 0 & \delta \leq B_{n,l}, \\ \delta^{-1}\,\tilde{\zeta}_\lambda(l, n, \delta, \hat{z}) > 0 & \delta > B_{n,l}. \end{cases}$$

Proof. According to (75), to determine the value of $\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta)$, we need to find the best point $\big(p_l(Q), f_l(Q)\big)$ in the region $\mathcal{R}_{l,n,\lambda}$ such that $p_l(Q) \geq \delta$ and $f_l(Q)/p_l(Q)$ is minimum. Note that this best point must be on the lower boundary of $\mathcal{R}_{l,n,\lambda}$.

First, we consider the case $0 < \delta \leq B_{n,l}$. According to Proposition 5, the point $\big(h_{n+1}(l, n, \lambda), g_{n+1}(l, n, \lambda)\big) = (B_{n,l}, 0)$ is in the region $\mathcal{R}_{l,n,\lambda}$. Hence, there exists a $Q \in \mathcal{Q}_{n+1}$ such that $p_l(Q) \geq \delta$ and $f_l(Q)/p_l(Q) = 0$. This implies that $\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta) = 0$.

Second, we consider the case $B_{n,l} < \delta \leq 1$. Let $\big(p_l(Q_A), f_l(Q_A)\big)$ be the point on the lower boundary of $\mathcal{R}_{l,n,\lambda}$ that satisfies $p_l(Q_A) = \delta$. According to Proposition 5 and Lemmas 4 and 5, if $\big(p_l(Q_B), f_l(Q_B)\big)$ is a point on the lower boundary of $\mathcal{R}_{l,n,\lambda}$ that satisfies $p_l(Q_B) > \delta$, then we must have $f_l(Q_B)/p_l(Q_B) > f_l(Q_A)/p_l(Q_A)$. Therefore, $\big(p_l(Q_A), f_l(Q_A)\big)$ is the best point $\big(p_l(Q), f_l(Q)\big)$ in the region $\mathcal{R}_{l,n,\lambda}$ such that $p_l(Q) \geq \delta$ and $f_l(Q)/p_l(Q)$ is minimum. This implies that

$$\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta) = \frac{f_l(Q_A)}{p_l(Q_A)} = \frac{f_l(Q_A)}{\delta}. \tag{81}$$

Since $p_l(Q_A) = \delta$ and $\hat{z}$ is the largest integer z such that $h_z(l, n, \lambda) \geq \delta$, the point $\big(p_l(Q_A), f_l(Q_A)\big)$ is on the line segment connecting the following two points,

$$\big(h_{\hat{z}+1}(l, n, \lambda), g_{\hat{z}+1}(l, n, \lambda)\big), \quad \big(h_{\hat{z}}(l, n, \lambda), g_{\hat{z}}(l, n, \lambda)\big).$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

Using this fact, one can easily verify that $f_l(Q_A) = \tilde{\zeta}_\lambda(l, n, \delta, \hat{z})$. This together with (81) completes the proof of Theorem 3. □

In particular, when $\delta = h_z(l, n, \lambda)$ for $z \in \{l, l+1, \ldots, n+1\}$, Theorem 3 implies that

$$\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta = h_z(l, n, \lambda)) = \frac{g_z(l, n, \lambda)}{h_z(l, n, \lambda)}, \tag{82}$$

which is non-increasing in z. This is because $h_z(l, n, \lambda)$ decreases monotonically with z by Lemma 4, and $\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta)$ is non-decreasing in δ for $0 < \delta \le 1$ by Proposition 6.

As a counterpart of Proposition 1 in Section 3.3, the following proposition clarifies the monotonicities of $\overline{\epsilon}(T_l(\lambda), n, \delta)$. See Appendix K for a proof.

Proposition 6. *Suppose* $0 \le \lambda < 1$, $0 < \delta \le 1$, $l \in \mathbb{Z}^{\ge 0}$, *and* $n \in \mathbb{Z}^{\ge l+1}$. *Then,* $\overline{\epsilon}(T_l(\lambda), n, \delta)$ *is non-increasing in* δ *for* $0 < \delta \le 1$, *non-increasing in* n *for* $n \in \mathbb{Z}^{\ge l+1}$, *and non-decreasing in* l *for* $0 \le l \le n-1$.

In the following, in order to show the remaining unproved results of Theorems 1 and 2, we prepare two useful statements, which provide several tight bounds for $\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta)$. Corollary 4 follows from (82), Theorem 3, Lemma 4, and the fact that $\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta)$ is non-decreasing in δ by Proposition 6. Lemma 6 is proved in Appendix L.

Corollary 4. *Suppose* $0 < \lambda < 1$, $0 < \delta < 1$, $l \in \mathbb{Z}^{\ge 0}$, *and* $n \in \mathbb{Z}^{\ge l+1}$, *then we have*

$$\frac{g_{z^*+1}(l, n, \lambda)}{h_{z^*+1}(l, n, \lambda)} \le \overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta) \le \max\left\{0, \frac{g_{z_*}(l, n, \lambda)}{h_{z_*}(l, n, \lambda)}\right\},$$

where z^* *and* z_* *are shorthand for* $z^*(l, \delta, \lambda)$ *and* $z_*(l, \delta, \lambda)$ *defined in (9), respectively.*

Lemma 6. *Suppose* $0 < \lambda < 1$, $0 < \delta < 1$, $l \in \mathbb{Z}^{\ge 0}$, *and* $n \in \mathbb{Z}^{\ge l+1}$. *Then*

$$\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta) \le \max\left\{0, \frac{\lambda(n - z_* + 1)}{\lambda(n - z_* + 1) + z_* - l}\right\}, \tag{83}$$

$$\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta) \ge \frac{\lambda(n - z^*)}{\lambda(n - z^*) + z^* + 1}. \tag{84}$$

If in addition $\delta \le \frac{1}{2}$ *then*

$$\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta) \ge \frac{\lambda(n - z^*)}{\lambda(n - z^*) + z^* - l + 1 + \sqrt{\lambda l}} \ge 1 - \frac{z^* - l + 1 + \sqrt{\lambda l}}{\lambda n}. \tag{85}$$

9. Proof of Theorem 1

We shall first prove (30) and then prove (32).

Proof of (30). Since the case $s = 0$ follows from (63), we focus on the case $0 < s < 1$. In the following, we let $l = \lceil \nu s n \rceil$. The proof is composed of two steps.

Step 1: The aim of this step is to show that (87) holds when $0 < s < 1$.

When $0 < s < 1$, in the limit $n \to +\infty$ we have

$$z^* = \frac{l - \Phi^{-1}(\delta)\sqrt{\lambda l}}{\nu} + O(1) = sn - \frac{\Phi^{-1}(\delta)\sqrt{\lambda \nu s n}}{\nu} + O(1) < n, \tag{86}$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

where the first equality follows from Lemma 17. Hence, the relation $B_{n,l} < B_{z^*,l} \le \delta$ holds as $n \to +\infty$, and we have $\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta) > 0$ according to Theorem 3. This fact together with Corollary 4 implies that

$$\frac{g_{z^*+1}(\lceil \nu s n \rceil, n, \lambda)}{h_{z^*+1}(\lceil \nu s n \rceil, n, \lambda)} \le \overline{\epsilon}^{\mathrm{c}}(T_{\lceil \nu s n \rceil}(\lambda), n, \delta) \le \frac{g_{z_*}(\lceil \nu s n \rceil, n, \lambda)}{h_{z_*}(\lceil \nu s n \rceil, n, \lambda)} \tag{87}$$

for a sufficiently large integer n.

Step 2: The aim of this step is to prove (30) in the case $0 < s < 1$. As shown in Appendix M, in the limit $n \to \infty$, we have

$$\text{RHS of (87)} = 1 - s - C(\lambda, s, \delta)\sqrt{\frac{1}{n}} + O\left(\frac{1}{n}\right). \tag{88}$$

Using a similar reasoning for deriving (88), we can deduce that

$$\text{LHS of (87)} = 1 - s - C(\lambda, s, \delta)\sqrt{\frac{1}{n}} + O\left(\frac{1}{n}\right). \tag{89}$$

Then (87)–(89) together confirm (30). □

Proof of (32). In the following, we let $l = \lceil \nu s n \rceil$ and $\delta = \mathrm{e}^{-rn}$. The proof is composed of two steps.

Step 1: In this step we consider the case with $r > D(s\nu \| \nu)$. In this case, we have

$$\lim_{n\to\infty} \frac{\ln \delta}{\ln B_{n,l}} \overset{(a)}{\ge} \lim_{n\to\infty} \frac{\ln(\mathrm{e}^{-rn})}{\ln\left[\frac{1}{\mathrm{e}\sqrt{\lceil \nu s n \rceil}} \mathrm{e}^{-D(\lceil \nu s n \rceil / n \| \nu) n}\right]} = \frac{r}{D(\nu s \| \nu)} > 1,$$

where (a) follows from the reverse Chernoff bound given in Lemma 8. This fact together with Theorem 3 implies that $\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta = \mathrm{e}^{-rn}) = 0$ when n is sufficiently large, which confirms (32).

Step 2: In this step we consider the case with $0 < r \le D(s\nu \| \nu)$.

Step 2.1: The aim of this step is to show that

$$z^* = r t_{\mathrm{D}}\left(\frac{\nu s}{r}, \nu\right) n + o(n), \tag{90}$$

where $t_{\mathrm{D}}(\frac{\nu s}{r}, \nu)$ is defined in (12).

Equations (A.1) and (A.2) imply that

$$\frac{1}{\mathrm{e}\sqrt{\lceil \nu s n \rceil}} \mathrm{e}^{-z^* D(\frac{\lceil \nu s n \rceil}{z^*} \| \nu)} \le B_{z^*, \lceil \nu s n \rceil} \le \delta = \mathrm{e}^{-rn} \le B_{z_*, \lceil \nu s n \rceil} \le \mathrm{e}^{-z_* D(\frac{\lceil \nu s n \rceil}{z_*} \| \nu)}.$$

It follows that

$$r = \lim_{n\to\infty} \frac{z_*}{n} D\left(\frac{\nu s n}{z^*} \middle\| \nu\right) = \left(\lim_{n\to\infty} \frac{z_*}{n}\right) D\left(\frac{\nu s}{\lim_{n\to\infty}(z^*/n)} \middle\| \nu\right),$$

and, thus, $\lim_{n\to\infty}(z_*/n) = r t_{\mathrm{D}}(\nu s/r, \nu)$, which confirms (90).

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

Step 2.2: The aim of this step is to show (32) by using (90) and Lemma 6.

According to Lemma 6 we have

$$\frac{\lambda(n-z^*)}{\lambda(n-z^*)+z^*-l+1+\sqrt{\lambda l}} \le \overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta) \le \max\left\{0, \frac{\lambda(n-z_*+1)}{\lambda(n-z_*+1)+z_*-l}\right\}. \tag{91}$$

Then we have

$$\lim_{n\to\infty} \text{LHS of (91)} = \lim_{n\to\infty} \frac{\lambda\left(1-\frac{z^*}{n}\right)}{\lambda\left(1-\frac{z^*}{n}\right)+\frac{z^*}{n}+\frac{1-\lceil \nu s n\rceil+\sqrt{\lambda\lceil \nu s n\rceil}}{n}} \tag{92}$$
$$\overset{(a)}{=} \frac{\lambda(1-rt_{\mathrm{D}}(\frac{\nu s}{r}, \nu))}{\lambda\left(1-rt_{\mathrm{D}}\left(\frac{\nu s}{r}, \nu\right)\right)+rt_{\mathrm{D}}\left(\frac{\nu s}{r}, \nu\right)-\nu s} = 1-E_{\lambda,s}(r),$$

where (a) follows from (90). By a similar reasoning, we can deduce that

$$\lim_{n\to\infty} \frac{\lambda(n-z_*+1)}{\lambda(n-z_*+1)+z_*-l} = 1-E_{\lambda,s}(r) \ge 0 \quad \text{for all } 0<r\le D(s\nu\|\nu).$$

Hence, we have $\lim_{n\to\infty}$ RHS of (91) $= 1-E_{\lambda,s}(r)$. This fact together with (92) and (91) confirms (32) in the case $0<r\le D(s\nu\|\nu)$. □

10. Conclusion

We have studied one-sided hypothesis testing under random sampling without replacement. That is, when $n+1$ binary random variables $X_1, \ldots, X_{n+1}$ are subject to a permutation invariant distribution and n binary random variables $X_1, \ldots, X_n$ are observed, we have proposed randomized tests with a randomization parameter $\lambda > 0$ for the expectation of the $(n+1)$th random variable X_{n+1} under a given significance level $\delta > 0$. We have shown that our proposed randomized test significantly improves deterministic test under random sampling without replacement. For example, as shown in (30) and (28) of Theorem 1, the asymptotic upper confidence limit in our randomized test is much better than that in the deterministic test in sampling without replacement. This problem setting appears under a certain adversarial setting. We can expect that our results will be applied to various areas when the data are generated in an adversarial way. Although our analysis is limited to one-sided hypothesis testing on a binary random variable, in principle, it can be generalized to the case with general non-binary random variables. In this way, our results are expected to have wider applications.

Appendix A. Useful Lemmas

We present several useful lemmas with their derivations.

A.1. Binomial probability

To show several important statements, here we prepare several useful lemmas to characterize $B_{z,l}(p)$ and $\Delta_{z,l}$, which are defined in (7) and (8), respectively. If there is no danger of confusion, we shall use $B_{z,l}$ as shorthand for $B_{z,l}(\nu)$, where $\nu := 1-\lambda$ and λ is the randomization parameter introduced in (3).

Lemma 7 *(Lemma 3.2 of [48]) Suppose* $0 \le l \le z$ *and* $0 < p < 1$. *Then* $B_{z,l}(p)$ *is strictly increasing in l, strictly decreasing in z, and non-increasing in p.*

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

The Chernoff bound states that

$$B_{z,l} \le \mathrm{e}^{-zD(l/z\|\nu)} \quad \text{for all } l \le \nu z, \tag{A.1}$$

where $D(p\|q)$ is the relative entropy defined in (11). The following lemma provides a reverse Chernoff bound for $B_{z,l}$.

Lemma 8. *(Proposition 5.4 of [48].) Suppose l,z are positive integers and satisfy $l \le z-1$, then*

$$B_{z,l} \ge \frac{1}{\mathrm{e}\sqrt{l}} \mathrm{e}^{-zD(l/z\|\nu)}. \tag{A.2}$$

The following lemma is a simple corollary of the Berry–Esseen theorem [6]

Lemma 9. *Suppose $0 < \lambda < 1$ and integers $z \ge l \ge 1$. Then*

$$\left| B_{z,l} - \Phi\left(\frac{l - \nu z}{\sqrt{\nu\lambda z}} \right) \right| \le \frac{0.5(1-2\nu\lambda)}{\sqrt{\nu\lambda z}} \le \frac{0.5}{\sqrt{\nu\lambda l}},$$

where $\Phi(x)$ is the standard normal distribution function defined in (10).

Lemma 10. *Suppose $0 < \lambda < 1$ and integers $0 \le l \le z$. Then $B_{z,l}/B_{z+1,l}$ is non-decreasing in z, and satisfies*

$$\frac{z-l+1}{(z+1)\lambda} \le \frac{B_{z,l}}{B_{z+1,l}} \le \frac{1}{\lambda}. \tag{A.3}$$

In addition, if $l \le \nu z$ then

$$\frac{B_{z,l}}{B_{z+1,l}} \le \frac{z-l+1+\sqrt{\lambda l}}{(z+1)\lambda}. \tag{A.4}$$

Proof of Lemma 10. If $l = 0$ then $B_{z,l}/B_{z+1,l} = 1/\lambda$ is nondecreasing in z. If $l \ge 1$ then the monotonicity of $B_{z,l}/B_{z+1,l}$ with z follows from Lemma 3.4 of [48].

In addition, (A.3) follows from Lemma 3.4 of [48], and (A.4) follows from Proposition 4.10 of [48]. □

Lemma 11. *Suppose $0 < \lambda < 1$ and $z \ge l+1$; then $\Delta_{z,l}/\Delta_{z-1,l}$ is non-increasing in z and non-decreasing in l.*

Proof of Lemma 11. The equality

$$B_{z+1,l} = (1-\lambda)B_{z,l-1} + \lambda B_{z,l},$$

implies that

$$\begin{aligned} \Delta_{z,l} &= (1-\lambda)(B_{z,l} - B_{z,l-1}) = (1-\lambda)b_{z,l} \\ &= \begin{cases} 0 & l \ge z+1, \\ (1-\lambda)\binom{z}{l}(1-\lambda)^l \lambda^{z-l} & l \le z. \end{cases} \end{aligned} \tag{A.5}$$

When $z \ge l+1$, (A.5) implies that

$$\frac{\Delta_{z,l}}{\Delta_{z-1,l}} = \frac{\binom{z}{l}\lambda}{\binom{z-1}{l}} = \frac{z\lambda}{z-l},$$

so $\Delta_{z,l}/\Delta_{z-1,l}$ is non-increasing in z and non-decreasing in l. □

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

A.2. Relative entropy

To show that $t_{\mathrm{D}}(\gamma, x)$, $\epsilon_D(s, r)$, and $s_D(\epsilon, r)$ are well defined in Section 3.2, we prepare the following lemma.

Lemma 12. *The relative entropy $D(p\|q)$ defined in (11) is non-decreasing in q and non-increasing in p when $0 \le p \le q \le 1$.*

Proof of Lemma 12. When $0 \le p \le q \le 1$ we have

$$\frac{\partial D(p\|q)}{\partial q} = \frac{q-p}{q(1-q)} \ge 0, \qquad \frac{\partial D(p\|q)}{\partial p} = \ln\frac{p}{q} + \ln\frac{1-q}{1-p} \le 0,$$

which confirm the lemma. □

A.3. Poisson distribution

To show that $t_{\mathrm{P}}(k, \delta)$ is well defined in (15), we prepare the following lemma.

Lemma 13. *Suppose $k \in \mathbb{Z}^{\ge 0}$. Then* $\mathrm{Pois}(k, x)$ *defined in (14) is monotonically decreasing in x for $x \ge 0$.*

Proof of Lemma 13. For $x > 0$ we have

$$\begin{aligned}\frac{\partial\,\mathrm{Pois}(k, x)}{\partial x} &= -\mathrm{e}^{-x}\sum_{i=0}^{k}\frac{x^i}{i!} + \mathrm{e}^{-x}\sum_{i=1}^{k} i\frac{x^{i-1}}{i!} \\ &= -\mathrm{e}^{-x}\sum_{i=0}^{k}\frac{x^i}{i!} + \mathrm{e}^{-x}\sum_{i=0}^{k-1}\frac{x^i}{i!} = -\mathrm{e}^{-x}\frac{x^k}{k!} < 0,\end{aligned}$$

which confirms the lemma. □

Appendix B. Properties of z^* and $B_{z_*,l}/B_{z^*,l}$

Given $0 < \delta \le 1$, $z^*(l, \delta, \lambda)$ is the smallest integer z that satisfies $B_{z,l} \le \delta$, as defined in (9). If there is no danger of confusion, we shall use z^* as shorthand for $z^*(l, \delta, \lambda)$, and use $z_* = z^* - 1$. This appendix provides some properties of z^* and $B_{z_*,l}/B_{z^*,l}$, which are useful for proving several important statements.

B.1. Lower bound for z^*

Lemma 14. *(Corollary 1 in [25].) Suppose $z \in \mathbb{Z}^{\ge 0}$ and $0 \le p \le 1$, then $B_{z,\lceil zp\rceil}(p) \ge \frac{1}{2}$.*

Lemma 15. *(Theorem 4.1 in [34].) Suppose $l \in \mathbb{Z}^{\ge 0}$ and $z \in \mathbb{Z}^{\ge l+1}$, then $B_{z,l}(l/z) \ne \frac{1}{2}$.*

Corollary 5. *Suppose $l \in \mathbb{Z}^{\ge 1}$ and $z \in \mathbb{Z}^{\ge l}$, then $B_{z,l}(l/z) > \frac{1}{2}$.*

Proof. First, in the case $z = l$, we have $B_{z,l}(l/z) = B_{z,z}(1) = 1 \ge \frac{1}{2}$. Second, in the case $z \ge l+1$, the corollary immediately follows from Lemmas 14 and 15. □

Corollary 6. *Suppose $0 < \lambda < 1$ and $0 < \delta \le \frac{1}{2}$. Then $z^* \ge l/\nu$.*

Proof. When $l = 0$, the lemma clearly holds.

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

Next, we shall consider $l \in \mathbb{Z}^{\geq 1}$. In this case we have $\lfloor l/\nu \rfloor \geq l$ and

$$B_{\lfloor l/\nu \rfloor, l}(\nu) \overset{(a)}{\geq} B_{\lfloor l/\nu \rfloor, l}\left(\frac{l}{\lfloor l/\nu \rfloor}\right) \overset{(b)}{>} \frac{1}{2}, \tag{B.1}$$

where (a) follows from Lemma 7 and the inequality $\nu \leq l/\lfloor l/\nu \rfloor$, and (b) follows from Corollary 5. Therefore,

$$z^* = \min\{z \in \mathbb{Z}^{\geq l} | B_{z,l}(\nu) \leq \delta\} \overset{(a)}{\geq} \min\left\{z \in \mathbb{Z}^{\geq l} | B_{z,l}(\nu) \leq \frac{1}{2}\right\} \overset{(b)}{\geq} \lfloor l/\nu \rfloor + 1 \geq l/\nu,$$

which confirms the lemma. Here (a) follows from the assumption $\delta \leq \frac{1}{2}$, and (b) follows from (B.1). □

B.2. Limits of z^*

Lemma 16. *Suppose $l \in \mathbb{Z}^{\geq 0}$, $0 < \lambda < 1$, and $0 < \delta < 1$, then*

$$\lim_{\delta \to 0} \frac{z^*}{\ln \delta} = \frac{1}{\ln \lambda}.$$

Proof of Lemma 16. By the large deviation principle, we have

$$\lim_{z \to \infty} \frac{1}{z} \ln (B_{z,l}) = -D(0 \| 1 - \lambda) = \ln \lambda.$$

Since we have $z^* \to \infty$ when $\delta \to 0$, the above relation implies that

$$\lim_{\delta \to 0} \frac{\ln \delta}{z^*} = \lim_{\delta \to 0} \frac{\ln (B_{z^*,l})}{z^*} = \lim_{z^* \to \infty} \frac{\ln (B_{z^*,l})}{z^*} = \ln \lambda,$$

which confirms the lemma. □

Lemma 17. *Suppose $l \in \mathbb{Z}^{\geq 0}$, $0 < \lambda < 1$, and $0 < \delta < 1$. Then in the limit $l \to +\infty$ we have*

$$z^* = \frac{l - \Phi^{-1}(\delta)\sqrt{\lambda l}}{\nu} + O(1),$$

where z^* is the smallest integer z such that $B_{z,l} \leq \delta$, and $\Phi(x)$ is defined in (10).

Proof of Lemma 17. By Lemma 9 we have

$$\Phi\left(\frac{l - \nu z}{\sqrt{\nu \lambda z}}\right) - \frac{0.5}{\sqrt{\nu \lambda l}} \leq B_{z,l} \leq \Phi\left(\frac{l - \nu z}{\sqrt{\nu \lambda z}}\right) + \frac{0.5}{\sqrt{\nu \lambda l}}. \tag{B.2}$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

When $l \to \infty$, we have $\delta - 0.5/\sqrt{\nu\lambda l} > 0$, and

$$\begin{aligned}
z^* &= \min\{z \in \mathbb{Z}^{\geq l} | B_{z,l} \leq \delta\} \\
&\leq \left\lceil \left\{ z \in \mathbb{R}^+ \;\middle|\; \Phi\left(\frac{l - \nu z}{\sqrt{\nu\lambda z}}\right) + \frac{0.5}{\sqrt{\nu\lambda l}} = \delta \right\} \right\rceil \\
&= \left\lceil \left\{ z \in \mathbb{R}^+ \;\middle|\; \frac{l - \nu z}{\sqrt{\nu\lambda z}} = \Phi^{-1}\left(\delta - \frac{0.5}{\sqrt{\nu\lambda l}}\right) \right\} \right\rceil \\
&= \left\lceil \left(-\sqrt{\frac{\lambda}{4\nu}}\, \Phi^{-1}\left(\delta - \frac{0.5}{\sqrt{\nu\lambda l}}\right) + \sqrt{\frac{l}{\nu} + \frac{\lambda}{4\nu}\left[\Phi^{-1}\left(\delta - \frac{0.5}{\sqrt{\nu\lambda l}}\right)\right]^2} \right)^2 \right\rceil \\
&= \left\lceil \left(-\sqrt{\frac{\lambda}{4\nu}}\, \Phi^{-1}(\delta) + O\left(\frac{1}{\sqrt{l}}\right) + \sqrt{\frac{l}{\nu} + O(1)} \right)^2 \right\rceil \\
&= \frac{l - \Phi^{-1}(\delta)\sqrt{\lambda l}}{\nu} + O(1),
\end{aligned} \tag{B.3}$$

where the inequality follows from the upper bound in (B.2).

In addition, when $l \to \infty$, we have $\delta + \frac{0.5}{\sqrt{\nu\lambda l}} < 1$, and

$$\begin{aligned}
z^* &= \min\{z \in \mathbb{Z}^{\geq l} | B_{z,l} \leq \delta\} \\
&\geq \left\{ z \in \mathbb{R}^+ \;\middle|\; \Phi\left(\frac{l - \nu z}{\sqrt{\nu\lambda z}}\right) - \frac{0.5}{\sqrt{\nu\lambda l}} = \delta \right\} \\
&= \left\{ z \in \mathbb{R}^+ \;\middle|\; \frac{l - \nu z}{\sqrt{\nu\lambda z}} = \Phi^{-1}\left(\delta + \frac{0.5}{\sqrt{\nu\lambda l}}\right) \right\} \\
&= \left(-\sqrt{\frac{\lambda}{4\nu}}\, \Phi^{-1}\left(\delta + \frac{0.5}{\sqrt{\nu\lambda l}}\right) + \sqrt{\frac{l}{\nu} + \frac{\lambda}{4\nu}\left[\Phi^{-1}\left(\delta + \frac{0.5}{\sqrt{\nu\lambda l}}\right)\right]^2} \right)^2 \\
&= \left(-\sqrt{\frac{\lambda}{4\nu}}\, \Phi^{-1}(\delta) + O\left(\frac{1}{\sqrt{l}}\right) + \sqrt{\frac{l}{\nu} + O(1)} \right)^2 \\
&= \frac{l - \Phi^{-1}(\delta)\sqrt{\lambda l}}{\nu} + O(1),
\end{aligned} \tag{B.4}$$

where the inequality follows from the lower bound in (B.2). Equations (B.3) and (B.4) together confirm the lemma. □

B.3. Limit of $B_{z_*,l}/B_{z^*,l}$

Lemma 18. *Suppose $l \in \mathbb{Z}^{\geq 0}$, $0 < \lambda < 1$ and $0 < \delta < 1$. Then in the limit $l \to +\infty$ we have*

$$\frac{B_{z_*,l}}{B_{z^*,l}} = 1 + \frac{\phi(\Phi^{-1}(\delta))\nu}{\delta\sqrt{\lambda l}} + O(l^{-1}), \tag{B.5}$$

$$\frac{B_{z_*-1,l}}{B_{z_*,l}} = 1 + \frac{\phi(\Phi^{-1}(\delta))\nu}{\delta\sqrt{\lambda l}} + O(l^{-1}), \tag{B.6}$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

$$\frac{B_{z^*,l}}{B_{z^*+1,l}} = 1 + \frac{\phi(\Phi^{-1}(\delta))\nu}{\delta\sqrt{\lambda l}} + O(l^{-1}). \tag{B.7}$$

Proof of Lemma 18. To consider the normal approximation, we define

$$x_1 := \frac{l - z^*\nu}{\sqrt{z^*\lambda\nu}}, \qquad x_2 := \frac{l - z_*\nu}{\sqrt{z_*\lambda\nu}}.$$

By Lemma 17, in the limit $l \to +\infty$ we have $z^* = l/\nu + O(\sqrt{l})$ and, thus,

$$x_1 = \frac{l - z^*\nu}{\sqrt{z^*\lambda\nu}} = \frac{l - \nu[l/\nu + O(\sqrt{l})]}{\sqrt{\lambda\nu[l/\nu + O(\sqrt{l})]}} = \frac{O(\sqrt{l})}{\sqrt{\lambda l + O(\sqrt{l})}} = O(1).$$

In addition, we have

$$\begin{aligned} x_2 - x_1 = \frac{l - z_*\nu}{\sqrt{z_*\lambda\nu}} - x_1 &= \frac{1}{\sqrt{z_*}}\sqrt{\frac{\nu}{\lambda}} + \frac{\sqrt{z^*}}{\sqrt{z_*}}\left(\frac{l - z^*\nu}{\sqrt{z^*\lambda\nu}}\right) - x_1 (B.9) \\ &= \frac{1}{\sqrt{z_*}}\sqrt{\frac{\nu}{\lambda}} + \left(\sqrt{1 + \frac{1}{z_*}} - 1\right) x_1 \\ &\overset{(a)}{=} \frac{1}{\sqrt{\frac{l}{\nu} + O\left(\sqrt{l}\right)}}\sqrt{\frac{\nu}{\lambda}} + \left(\sqrt{1 + O\left(\frac{1}{l}\right)} - 1\right) x_1 \\ &\overset{(b)}{=} \left[1 + O\Big(\frac{1}{\sqrt{l}}\Big)\right] \frac{\nu}{\sqrt{\lambda l}} + O(l^{-1}) = \frac{\nu}{\sqrt{\lambda l}} + O(l^{-1}), \end{aligned} \tag{B.8}$$

where (a) holds because $z_* = l/\nu + O(\sqrt{l})$ and $1/z_* = O(1/l)$ (see Lemma 17), and (b) holds because $x_1 = O(1)$.

According to Lemma 9 we have

$$B_{z^*,l} = \Phi\left(\frac{l - z^*\nu}{\sqrt{z^*\lambda\nu}}\right) + O\Big(\frac{1}{\sqrt{l}}\Big) = \Phi(x_1) + O\Big(\frac{1}{\sqrt{l}}\Big). \tag{B.9}$$

In addition, considering the Edgeworth series [21, (2.1)], we have

$$\frac{B_{z_*,l}}{B_{z^*,l}} = \frac{\Phi(x_2)}{\Phi(x_1)} + O(l^{-1}) = \frac{\Phi(x_1) + \phi(x_3)(x_2 - x_1)}{\Phi(x_1)} + O(l^{-1})$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

where $x_3 \in (\min(x_1, x_2), \max(x_1, x_2))$. By (B.8), we have $x_2 - x_1 = \nu/\sqrt{\lambda l} + O(l^{-1})$ and $x_3 = x_1 + O(1/\sqrt{l})$. Hence,

$$\begin{aligned}\frac{B_{z_*,l}}{B_{z^*,l}} &= \frac{\Phi(x_1) + \phi(x_3)(x_2 - x_1)}{\Phi(x_1)} + O(l^{-1}) \\ &= 1 + \frac{\phi\left(x_1 + O(\frac{1}{\sqrt{l}})\right)\left(\frac{\nu}{\sqrt{\lambda l}} + O(l^{-1})\right)}{\Phi(x_1)} + O(l^{-1}) \\ &= 1 + \frac{\left[\phi(x_1) + O(\frac{1}{\sqrt{l}})\right]\left(\frac{\nu}{\sqrt{\lambda l}} + O(l^{-1})\right)}{\Phi(x_1)} + O(l^{-1}) \\ &= 1 + \frac{\phi(x_1)\nu}{\Phi(x_1)\sqrt{\lambda l}} + O(l^{-1}), \end{aligned} \tag{B.10}$$

Since $B_{z^*,l} \le \delta < B_{z_*,l}$, (B.10) implies that $B_{z^*,l} = \delta + O\left(\frac{1}{\sqrt{l}}\right)$. This fact together with (B.9) implies that

$$\Phi(x_1) = \delta + O\left(\frac{1}{\sqrt{l}}\right), \qquad x_1 = \Phi^{-1}(\delta) + O\left(\frac{1}{\sqrt{l}}\right).$$

Then by plugging the above two equations into (B.10) we get (B.5).

Equations (B.6) and (B.7) follow from (B.5) and the relations $B_{z^*,l} = \delta + O\left(\frac{1}{\sqrt{l}}\right)$ and $B_{z_*,l} = \delta + O\left(\frac{1}{\sqrt{l}}\right)$. □

Appendix C. Proof of Proposition 1

According to (6) we have

$$\overline{\epsilon}_{\text{i.i.d.}}(T_k, n, \delta) = \max\{0 \le \epsilon \le 1 \,|B_{n,k}(\epsilon) \ge \delta\},$$

where $B_{n,k}(\epsilon)$ is defined in (7).
Monotonicity in δ: The monotonicity of $\overline{\epsilon}_{\text{i.i.d.}}(T_k, n, \delta)$ with δ follows immediately from (75).
Monotonicity in n: We have

$$\begin{aligned}\overline{\epsilon}_{\text{i.i.d.}}(T_k, n, \delta) &= \max\{0 \le \epsilon \le 1 \,|B_{n,k}(\epsilon) \ge \delta\} \\ &\ge \max\{0 \le \epsilon \le 1 \,|B_{n+1,k}(\epsilon) \ge \delta\} = \overline{\epsilon}_{\text{i.i.d.}}(T_k, n+1, \delta),\end{aligned}$$

where the inequality holds because $B_{n,k}(\epsilon) \ge B_{n+1,k}(\epsilon)$ (according to Lemma 7). Therefore, $\overline{\epsilon}_{\text{i.i.d.}}(T_k, n, \delta)$ is non-increasing in n for $n \ge k+1$.
Monotonicity in k: We have

$$\begin{aligned}\overline{\epsilon}_{\text{i.i.d.}}(T_k, n, \delta) &= \max\{0 \le \epsilon \le 1 \,|B_{n,k}(\epsilon) \ge \delta\} \\ &\le \max\{0 \le \epsilon \le 1 \,|B_{n,k+1}(\epsilon) \ge \delta\} = \overline{\epsilon}_{\text{i.i.d.}}(T_{k+1}, n, \delta),\end{aligned}$$

where the inequality holds because $B_{n,k}(\epsilon) \le B_{n,k+1}(\epsilon)$ (according to Lemma 7). Therefore, $\overline{\epsilon}_{\text{i.i.d.}}(T_k, n, \delta)$ is nondecreasing in k for $0 \le k \le n-1$.

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

Appendix D. Proof of Lemma 1

Proof of (16). First, we show (16) by showing the relation

$$\max\left\{0 \le \epsilon \le 1 \,|B_{n,k}(\epsilon) \ge \frac{1}{2}\right\} > \frac{k}{n} \tag{D.1}$$

because (6) guarantees the equation $\max\{0 \le \epsilon \le 1 \,|B_{n,k}(\epsilon) \ge \delta\} = \overline{\epsilon}_{\text{i.i.d.}}(T_k, n, \delta)$ and the relation $\max\{0 \le \epsilon \le 1 \,|B_{n,k}(\epsilon) \ge \delta\} \ge \max\{0 \le \epsilon \le 1 \,|B_{n,k}(\epsilon) \ge \frac{1}{2}\}$ holds for $0 < \delta \le \frac{1}{2}$. For this aim, we shall consider two cases depending on the value of k.

(1) $k = 0$. In this case we have

$$\max\left\{0 \le \epsilon \le 1 \,|B_{n,k}(\epsilon) \ge \frac{1}{2}\right\} = \max\left\{0 \le \epsilon \le 1 \,|(1-\epsilon)^n \ge \frac{1}{2}\right\} > 0,$$

which confirms (D.1).

(2) $k \in \mathbb{Z}^{\ge 1}$. In this case we have $B_{n,k}(k/n) > \frac{1}{2}$ according to Corollary 5. Since $B_{n,k}(\epsilon)$ is continuous in ϵ, there exists a positive number $\epsilon_0 > k/n$ such that $B_{n,k}(\epsilon_0) \ge \frac{1}{2}$. Hence, we find that

$$\max\left\{0 \le \epsilon \le 1 \,|B_{n,k}(\epsilon) \ge \tfrac{1}{2}\right\} \ge \epsilon_0 > \frac{k}{n},$$

which implies (D.1).

In conclusion, (D.1) holds for all $k \in \mathbb{Z}^{\ge 0}$. □

Proof of (17). We show (17) by using (16). When $0 < \delta \le \frac{1}{2}$ we have

$$\begin{aligned}\bar{\epsilon}(T_l(\lambda), n, \delta) \ge \bar{\epsilon}\left(T_l(\lambda), n, \frac{1}{2}\right) &\overset{(a)}{\ge} \max_{Q \in \mathcal{Q}_{\text{i.i.d.},n+1}} \left\{Q\left(Y_{n+1} = 1 | L \le l\right) \mid Q(L \le l) \ge \frac{1}{2}\right\} \\ &\overset{(b)}{=} \max\left\{0 \le \epsilon \le 1 \,|B_{n,l}(\nu\epsilon) \ge \frac{1}{2}\right\}.\end{aligned} \tag{D.2}$$

Here (a) holds because the optimization constraints on the right-hand side are stronger than those in $\bar{\epsilon}_\lambda(l, n, \frac{1}{2})$; (b) follows from the fact that $\mathcal{Q}_{\text{i.i.d.},n+1} = \{P_\theta^{n+1}\}_{\theta \in [0,1]}$.

In the following, we consider two cases depending on the relation between νn and l. First, when $\nu n \le l$, we have

$$B_{n,l}(\nu) \overset{(a)}{\ge} B_{n,l}(l/n) \overset{(b)}{\ge} \frac{1}{2},$$

where (a) follows from $l/n \ge \nu$ and Lemma 7; and (b) follows from Lemma 14. This equation together with (D.2) implies that $\bar{\epsilon}_\lambda(l, n, \frac{1}{2}) = 1$, which confirms (17).

Second, when $\nu n > l$, we have

$$\text{RHS of (D.2)} = \frac{1}{\nu} \max\left\{0 \le \epsilon \le \nu \,|B_{n,l}(\epsilon) \ge \frac{1}{2}\right\} \overset{(a)}{>} \frac{l}{\nu n},$$

which confirms (17) again. Here (a) follows from (16), i.e. (D.1). □

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

Proof of (18). When $k/n < \delta < (k+1)/(n+1)$, (1) implies (18). When $(k+1)/(n+1) \le \delta \le \frac{1}{2}$, we have

$$\begin{aligned}\overline{\epsilon}(T_k(0), n, \delta) &= \frac{(k+1)(n+1-k) - \delta(n+1)}{\delta(n-k)(n+1)} \ge \frac{(k+1)(n+1-k) - \frac{1}{2}(n+1)}{\delta(n-k)(n+1)} \\ &= \frac{k(n-k) + \frac{1}{2}(n+1)}{\delta(n-k)(n+1)} > \frac{k}{\delta n},\end{aligned}$$

which confirms (18) again. Here the last inequality holds because the following conditions are equivalent,

$$\begin{aligned}\frac{k(n-k) + \frac{1}{2}(n+1)}{\delta(n-k)(n+1)} > \frac{k}{\delta n} \quad &\Leftrightarrow \quad nk(n-k) + \frac{n}{2}(n+1) > k(n-k)(n+1) \\ &\Leftrightarrow \quad \frac{n}{2}(n+1) > k(n-k), \end{aligned} \tag{D.3}$$

and that (D.3) holds for $(k+1)/(n+1) \le \frac{1}{2}$. □

Appendix E. Proof of Proposition 2

Recall that $\mathcal{Q}_{\text{i.i.d.},n+1} = \{P_\theta^{n+1}\}_{\theta\in[0,1]}$, where P_θ^{n+1} is the i.i.d. distribution on $[n+1]$ with $\Pr(Y_i = 1) = \theta$. The central limit theorem states that

$$\lim_{n\to\infty} P_\theta^{n+1}\left(K \le n\left(\theta + \frac{z}{\sqrt{n}}\right)\right) = \Phi\left(\frac{z}{\sqrt{\mathrm{Var}[Y_i]}}\right) = \Phi\left(\frac{z}{\sqrt{\theta(1-\theta)}}\right)$$

for any real number z, where $\mathrm{Var}[Y_i]$ is the variance of Y_i. As a consequence, we have

$$\lim_{n\to\infty} P_{s+\frac{\tau}{\sqrt{n}}}^{n+1}(K \le \lceil sn\rceil) = \lim_{n\to\infty} \Phi\left(\frac{\left(\frac{\lceil sn\rceil}{n} - s - \frac{\tau}{\sqrt{n}}\right)\sqrt{n}}{\sqrt{\left(s+\frac{\tau}{\sqrt{n}}\right)\left(1-s-\frac{\tau}{\sqrt{n}}\right)}}\right) = \Phi\left(-\frac{\tau}{\sqrt{s(1-s)}}\right).$$

Hence, the condition $\lim_{n\to\infty} P_{s+\frac{\tau}{\sqrt{n}}}^{n+1}(K \le \lceil sn\rceil) \le \delta$ is equivalent to the condition $\tau \le -\Phi^{-1}(\delta)\sqrt{s(1-s)}$. The above relation shows case (a) for $\overline{\epsilon}_{\text{i.i.d.}}(T_{\lceil sn\rceil}, n, \delta)$.

In addition, we have

$$\lim_{n\to\infty} \frac{-1}{n} \ln P_\tau^{n+1}(K \le \lceil sn\rceil) = D(s\|\tau).$$

Solving the equation $D(s\|\tau) = r$ for τ, we have $\tau = \epsilon_D(s, r)$. The above relation shows case (b) for $\overline{\epsilon}_{\text{i.i.d.}}(\lceil sn\rceil, n, \delta)$.

Appendix F. Proof of Lemma 2

First, the derivative of $\kappa(\lambda)$ reads

$$\begin{aligned}\kappa'(\lambda) &= \frac{((1-\theta_0)\alpha - (1+\theta_0))\beta\lambda^{-1/2} - (1-\theta_0)\beta\lambda^{-3/2} - \gamma(1-\lambda)^{-1/2}}{2(1-\theta_0(1-\lambda))^{3/2}} \\ &= \frac{\Phi^{-1}(\delta)(1-\theta_0)\left((1+(1+\theta_0)\psi(\delta))\lambda^{-1/2} + (1-\theta_0)\psi(\delta)\lambda^{-3/2}\right) - \frac{e-t}{\sqrt{\theta_0}}(1-\lambda)^{-1/2}}{2(1-\theta_0(1-\lambda))^{3/2}},\end{aligned}$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

which confirms the equivalence between $\kappa'(\lambda)=0$ and (58). We note that (58) is equivalent to the cubic equation $f_1(\lambda)+f_2(\lambda)=0$, where the functions $f_1(x)$ and $f_2(x)$ are defined as follows,

$$\begin{aligned} f_1(x) &:= \frac{(e-t)^2}{\theta(\Phi^{-1}(\delta)(1-\theta_0))^2}x^3, \\ f_2(x) &:= (x-1)\left[(1+(1+\theta_0)\psi(\delta))x+(1-\theta_0)\psi(\delta)\right]^2. \end{aligned}$$

The cubic equation $f_2(x)=0$ has two real roots $x_0=\frac{-(1-\theta_0)\psi(\delta)}{1+(1+\theta_0)\psi(\delta)}$ and $x_1=1$, where x_0 is a double root.

The remaining proof is composed of three steps. The results of steps 1–3 will together confirm the fact that (58) has but only one solution in (0, 1).

Step 1: The aim of this step is to show that the equation $f_1(x)+f_2(x)=0$ has at least one solution in (0, 1).

We have

$$\begin{aligned} f_1(0)+f_2(0) &= -\left[(1-\theta_0)\psi(\delta)\right]^2<0, \\ f_1(1)+f_2(1) &= \frac{(e-t)^2}{\theta(\Phi^{-1}(\delta)(1-\theta_0))^2}>0. \end{aligned} \tag{F.1}$$

Since $f_1(x)+f_2(x)$ is continuous in x, the above inequalities imply that the equation $f_1(x)+f_2(x)=0$ has at least one solution in (0, 1).

Step 2: The aim of this step is to show the double root $x_0<0$.

First note that the numerator satisfies $-(1-\theta_0)\psi(\delta)>0$ for $\delta<\frac{1}{2}$. Second, the denominator satisfies

$$1+(1+\theta_0)\psi(\delta)\overset{(a)}{<}1+\psi(\delta)=1+\frac{\phi(\Phi^{-1}(\delta))}{\delta\Phi^{-1}(\delta)}\overset{(b)}{=}1-\frac{\phi(y)}{y\Phi(-y)}\overset{(c)}{\leq}0,$$

which confirms that $x_0<0$. Here (a) holds because $\psi(\delta)<0$ when $\delta<\frac{1}{2}$; in (b) we change the variable by letting $y=-\Phi^{-1}(\delta)$; and (c) holds because $\Phi(-y)\leq\phi(y)/y$ for $y>0$, witch was first proved by Gordon [17].

Step 3: The aim of this step is to show that the equation $f_1(x)+f_2(x)=0$ has at most one solution in $(0,\infty)$. Since the double root $x_0<0$, there are only two cases for the monotonicity of $f_2(x)$ in $(0,\infty)$:

Case A: $f_2(x)$ increases monotonically with x for $x\in(0,\infty)$. In this case, $f_1(x)+f_2(x)$ is non-decreasing in x for $x\in(0,\infty)$. Hence, the equation $f_1(x)+f_2(x)=0$ has at most one solution in $(0,\infty)$.

Case B: In $(0,\infty)$, $f_2(x)$ first decreases monotonically and then increases monotonically. In this case, in the interval $(0,\infty)$, the cubic function $f_1(x)+f_2(x)$ also first monotonically decreases and then monotonically increases. This fact together with the first inequality in (F.1) implies that the equation $f_1(x)+f_2(x)=0$ has at most one solution in $(0,\infty)$.

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

Appendix G. Proof of Proposition 4

Remember that $\mathcal{Q}_{\text{i.i.d.},n+1} = \{P_\theta^{n+1}\}_{\theta\in[0,1]}$. The law of small numbers [37] guarantees that

$$\lim_{n\to\infty} P_{\frac{\tau}{n}}^{n+1}(K \le k_0) = \text{Pois}(k_0, \tau).$$

Hence, the condition $\text{Pois}(k_0, \tau) = \lim_{n\to\infty} P_{\tau/n}^{n+1}(K \le k_0) \le \delta$ is equivalent to the condition $\tau \le t_{\text{P}}(k_0, \delta)$. The above relation shows Case (a) for $\overline{\epsilon}_{\text{i.i.d.}}(T_{k_0}, n, \delta)$.

The large deviation principle [13] guarantees that

$$\lim_{n\to\infty} \frac{-1}{n} \ln P_\tau^{n+1}(K \le k_0) = D(0\|\tau) = -\ln(1-\tau).$$

Solving the equation $-\ln(1-\tau) = r$ for τ, we have $\tau = 1 - \mathrm{e}^{-r}$. The above relation shows case (b) for $\overline{\epsilon}_{\text{i.i.d.}}(T_{k_0}, n, \delta)$.

Appendix H. Proof of Lemma 3

Proof of Lemma 3. First, by (72) we have

$$\lim_{\epsilon\to 0} \lim_{\delta\to 0} \frac{\epsilon N_{\lambda,\text{c}}(\lceil \nu k_0 \rceil, \epsilon, \delta)}{\ln \delta^{-1}} = \frac{1}{\lambda \ln \lambda^{-1}}. \tag{H.1}$$

In the following, we shall prove that

$$\lim_{\delta\to 0} \lim_{\epsilon\to 0} \frac{\epsilon N_{\lambda,\text{c}}(\lceil \nu k_0 \rceil, \epsilon, \delta)}{\ln \delta^{-1}} = \frac{1}{\lambda \ln \lambda^{-1}}, \tag{H.2}$$

which together with (H.1) confirms the lemma.

For convenience, we use $l = \lceil \nu k_0 \rceil$. By Lemma 19 below, when $\delta \le \frac{1}{2}$ we have

$$\frac{(1-\epsilon)(z_* - l)}{\lambda \ln \delta^{-1}} + \frac{\epsilon(z_* - 1)}{\ln \delta^{-1}} \le \frac{\epsilon N_{\lambda,\text{c}}(l, \epsilon, \delta)}{\ln \delta^{-1}} \le \frac{z^* - l + 1 + \sqrt{\lambda l}}{\lambda \ln \delta^{-1}}, \tag{H.3}$$

where the upper bound follows from (H.5), and the lower bound follows from (H.4). Then Lemma 16 implies that

$$\lim_{\delta\to 0} \lim_{\epsilon\to 0} \text{LHS of (H.3)} = \lim_{\delta\to 0} \frac{z^* - l - 1}{\lambda \ln \delta^{-1}} = \frac{1}{\lambda \ln \lambda^{-1}},$$
$$\lim_{\delta\to 0} \lim_{\epsilon\to 0} \text{RHS of (H.3)} = \lim_{\delta\to 0} \frac{z^* - l + 1 + \sqrt{\lambda l}}{\lambda \ln \delta^{-1}} = \frac{1}{\lambda \ln \lambda^{-1}},$$

which confirm (H.2) and complete the proof. □

Lemma 19. *Suppose* $0 < \lambda, \epsilon, \delta < 1$ *and* $l \in \mathbb{Z}^{\ge 0}$. *Then*

$$N_{\lambda,\text{c}}(l, \epsilon, \delta) \ge \left(\frac{1}{\epsilon} - 1\right) \frac{z_* B_{z_*-1,l}}{B_{z_*,l}} + z_* - 1 \ge \left(\frac{1}{\epsilon} - 1\right) \frac{z_* - l}{\lambda} + z_* - 1. \tag{H.4}$$

If in addition $\delta \le \frac{1}{2}$ *then*

$$N_{\lambda,\text{c}}(l, \epsilon, \delta) \le \left\lceil \frac{z^* - l + 1 + \sqrt{\lambda l}}{\lambda} \left(\frac{1}{\epsilon} - 1\right) + z^* \right\rceil < \frac{z^* - l + 1 + \sqrt{\lambda l}}{\lambda \epsilon}. \tag{H.5}$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

Proof of Lemma 19. The proof is composed of two steps.

Step 1: Proof of (H.4). When $0 < \delta < 1$ we have

$$\begin{aligned} N_{\lambda,\mathrm{c}}(l,\epsilon,\delta) &= \min\{n \geq l+1 \mid \overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta) \geq 1-\epsilon\} \\ &\geq \min\left\{n \,\middle|\, \max\left\{0, \frac{g_{z_*}(l,n,\lambda)}{h_{z_*}(l,n,\lambda)}\right\} \geq 1-\epsilon\right\} \\ &= \left\lceil \frac{z_* B_{z_*-1,l}}{B_{z_*,l}}\left(\frac{1}{\epsilon}-1\right) + z_* - 1 \right\rceil \geq \frac{z_*-l}{\lambda}\left(\frac{1}{\epsilon}-1\right) + z_* - 1, \end{aligned}$$

which confirms (H.4). Here the first inequality follows from Corollary 4 in the main text, and the second equality follows from Lemma 10.

Step 2: Proof of (H.5). When $0 < \delta \leq \frac{1}{2}$ we have

$$\begin{aligned} N_{\lambda,\mathrm{c}}(l,\epsilon,\delta) &= \min\{n \geq l+1 \mid \overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta) \geq 1-\epsilon\} \\ &\leq \min\left\{n \geq l+1 \,\middle|\, \frac{g_{z^*+1}(l,n,\lambda)}{h_{z^*+1}(l,n,\lambda)} \geq 1-\epsilon\right\} = \left\lceil \frac{(z^*+1)B_{z^*,l}}{B_{z^*+1,l}}\left(\frac{1}{\epsilon}-1\right) + z^* \right\rceil \\ &\leq \left\lceil \frac{z^*-l+1+\sqrt{\lambda l}}{\lambda}\left(\frac{1}{\epsilon}-1\right) + z^* \right\rceil < \frac{z^*-l+1+\sqrt{\lambda l}}{\lambda\epsilon}, \end{aligned}$$

which confirms (H.5). Here the first inequality follows from Corollary 4 in the main text, and the second inequality follows from Lemma 10 and the fact that $z^* \geq l/\nu$ when $0 < \delta \leq \frac{1}{2}$ (see Corollary 6). □

Appendix I. Proofs of (76) and (77)

For $z \in \{0, 1, \ldots, n+1\}$, we have

$$\begin{aligned} Q(K = z, Y_{n+1} = 0 | Z = z) &= \frac{n-z+1}{n+1}, \\ Q(K = z-1, Y_{n+1} = 1 | Z = z) &= \frac{z}{n+1}. \end{aligned}$$

In the following, we shall show (76) and (77), respectively.

Step 1: Proof of (76). For $0 \leq v \leq z-1$, we have

$$\begin{aligned} Q(L = v | Z = z) &= Q(K = z, Y_{n+1} = 0 | Z = z)\binom{z}{v}(1-\lambda)^v \lambda^{z-v} \\ &\quad + Q(K = z-1, Y_{n+1} = 1 | Z = z)\binom{z-1}{v}(1-\lambda)^v \lambda^{z-1-v} \\ &= \frac{n-z+1}{n+1} b_{z,v}(v) + \frac{z}{n+1} b_{z-1,v}(v), \end{aligned}$$

where $b_{z-1,v}(v)$ is defined in (7). Hence, for $0 \leq l \leq z-1$, we have

$$\begin{aligned} Q(L \leq l | Z = z) &= \sum_{v=0}^{l} Q(L = v | Z = z) = \frac{n-z+1}{n+1}\sum_{v=0}^{l} b_{z,v}(v) + \frac{z}{n+1}\sum_{v=0}^{l} b_{z-1,v}(v) \\ &= \frac{n-z+1}{n+1} B_{z,l} + \frac{z}{n+1} B_{z-1,l}. \end{aligned}$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

In addition, when $l \geq z$, it is obvious that $P(L \leq l|Z = z) = 1$. This completes the proof of (76).

Step 2: Proof of (77). For $0 \leq v \leq z-1$, we have

$$Q(L = v, Y_{n+1} = 0|Z = z) = Q(K = z, Y_{n+1} = 0|Z = z)\binom{z}{v}(1-\lambda)^v \lambda^{z-v}$$
$$= \frac{n-z+1}{n+1} b_{z,v}(v).$$

Hence, when $0 \leq l \leq z-1$, we have

$$Q(L \leq l, Y_{n+1} = 0|Z = z) = \sum_{v=0}^{l} Q(L = v, Y_{n+1} = 0|Z = z)$$
$$= \frac{n-z+1}{n+1} \sum_{v=0}^{l} b_{z,v}(v) = \frac{n-z+1}{n+1} B_{z,l}.$$

In addition, when $l \geq z$, we have

$$Q(L \leq l, Y_{n+1} = 0|Z = z) = Q(K = z, Y_{n+1} = 0|Z = z) = \frac{n-z+1}{n+1}.$$

This completes the proof of (77).

Appendix J. Proofs of Lemmas 4 and 5

Proof of Lemma 4. According to Lemma 7, $B_{z,l}$ decreases monotonically in z for $z \geq l$. For $z \in \{l+1, \ldots, n\}$, we have

$$h_{n+1}(l, n, \lambda) = B_{n,l} \leq B_{z,l} < h_z(l, n, \lambda) < B_{z-1,l} \leq B_{l,l} = h_l(l, n, \lambda),$$

where the second and third inequality hold because

$$h_z(l, n, \lambda) = B_{z,l} + \frac{z}{n+1}(B_{z-1,l} - B_{z,l}).$$

Hence, $h_z(l, n, \lambda)$ decreases monotonically in z for $l \leq z \leq n+1$.

The monotonicity of $g_z(l, n, \lambda)$ with z follows from the facts that $(n-z+1)/(n+1)$ decreases monotonically in z for $0 \leq z \leq n+1$, and that $B_{z,l}$ decreases monotonically in z for $l \leq z \leq n+1$.

By (77), $g_z(l, n, \lambda) \geq 0$ for $z \leq n+1$, and $g_z(l, n, \lambda) < 0$ for $z \geq n+2$. In addition, By (76) we have

$$h_z(l, n, \lambda) = B_{z,l} + \frac{z}{n+1}(B_{z-1,l} - B_{z,l}) > 0,$$

which completes the proof. □

Proof of Lemma 5. Due to Lemma 4, in order to prove Lemma 5, it suffices to show that $[h_{z+1}(l, n, \lambda) - h_z(l, n, \lambda)]/[g_{z+1}(l, n, \lambda) - g_z(l, n, \lambda)]$ is strictly increasing in z for $l \leq z \leq n$. We have

$$(n+1)[h_{z+1}(l, n, \lambda) - h_z(l, n, \lambda)] = (n-z)B_{z+1,l} + (2z-n)B_{z,l} - zB_{z-1,l},$$
$$(n+1)[g_{z+1}(l, n, \lambda) - g_z(l, n, \lambda)] = (n-z)B_{z+1,l} - (n-z+1)B_{z,l},$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

which imply that

$$\frac{h_{z+1}(l,n,\lambda)-h_z(l,n,\lambda)}{g_{z+1}(l,n,\lambda)-g_z(l,n,\lambda)}-1=\frac{z(B_{z-1,l}-B_{z,l})-B_{z,l}}{(n-z)(B_{z,l}-B_{z+1,l})+B_{z,l}}$$
$$=\frac{z\Delta_{z-1,l}-B_{z,l}}{(n-z)\Delta_{z,l}+B_{z,l}}=\frac{z\Delta_{z-1,l}}{(n-z)\Delta_{z,l}+B_{z,l}}\left(1-\frac{B_{z,l}}{z\Delta_{z-1,l}}\right).$$

Note that

$$\frac{\Delta_{z-1,l}}{B_{z,l}}=\frac{B_{z-1,l}}{B_{z,l}}-1$$

is non-decreasing in z according to Lemma 10, so the function $\left(1-\frac{B_{z,l}}{z\Delta_{z-1,l}}\right)$ is strictly increasing in z. In addition, $\Delta_{z-1,l}/\Delta_{z,l}$ is nondecreasing in z according to Lemma 11, so the function $z\Delta_{z-1,l}/[(n-z)\Delta_{z,l}+B_{z,l}]$ is strictly increasing in z, because

$$\frac{(n-z)\Delta_{z,l}+B_{z,l}}{z\Delta_{z-1,l}}=\left(\frac{n}{z}-1\right)\frac{\Delta_{z,l}}{\Delta_{z-1,l}}+\frac{B_{z,l}}{z\Delta_{z-1,l}}$$

is strictly decreasing in z. Therefore, $[h_{z+1}(l,n,\lambda)-h_z(l,n,\lambda)]/[g_{z+1}(l,n,\lambda)-g_z(l,n,\lambda)]$ is strictly increasing in z for $l\le z\le n$, which completes the proof. □

Appendix K. Proof of Proposition 6

We show Proposition 6 by proving the following lemma.

Lemma 20. *Suppose* $0\le\lambda<1$, $0<\delta\le 1$, $l\in\mathbb{Z}^{\ge 0}$, *and* $n\in\mathbb{Z}^{\ge l+1}$. *Then* $\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda),n,\delta)$ *is non-decreasing in* δ *for* $0<\delta\le 1$, *non-decreasing in* n *for* $n\in\mathbb{Z}^{\ge l+1}$, *and non-increasing in* l *for* $0\le l\le n-1$.

Monotonicity in δ: The monotonicity of $\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda),n,\delta)$ with δ follows immediately from its definition in (75).

Monotonicity in n: To prove the monotonicity of $\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda),n,\delta)$ with n, we consider two cases depending on the value of λ. First, when $\lambda=0$, the monotonicity follows from (1). Second, when $\lambda>0$, we show the monotonicity by showing that the lower boundary of $\mathcal{R}_{n,l,\lambda}$ is not higher than that of $\mathcal{R}_{n+1,l,\lambda}$ for $n\ge l+1$. This fact can be shown as follows.

Since $g_z(l,n,\lambda)$ and $h_z(l,n,\lambda)$ are defined in (77) and (76), respectively, we have the relations

$$g_z(l,n,\lambda)\le g_z(l,n+1,\lambda),\quad h_z(l,n,\lambda)\ge h_z(l,n+1,\lambda)$$

for $z=0,1,\ldots,n+1$. Let $\tau_{n,l}(x)$ be the function that corresponds to the lower boundary of $\mathcal{R}_{n,l,\lambda}$. For $x\le h_{n+1}(l,n,\lambda)$, we define $\tau_{n,l}(x)$ to be 0. We denote by $\kappa[A_1,A_2]$ the line that passes through the two points A_1 and A_2, and by $A_z(l,n)$ the point $(h_z(l,n,\lambda),g_z(l,n,\lambda))$. Since Lemma 4 guarantees that $h_z(l,n,\lambda)$ is monotonically decreasing in z, for $x\le h_z(l,n,\lambda)$, there exists an integer $z'\ge z+1$ such that $x\in[h_{z'}(l,n,\lambda),h_{z'-1}(l,n,\lambda)]$. Then, we have

$$\tau_{n,l}(x)\le\kappa[A_z(l,n),A_{z'}(l,n)](x). \tag{K.1}$$

Since Lemma 4 guarantees that $g_z(l,n,\lambda)$ is monotonically decreasing in z, we have

$$g_z(l,n+1,\lambda)\ge g_z(l,n,\lambda)\ge g_{z'}(l,n,\lambda). \tag{K.2}$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

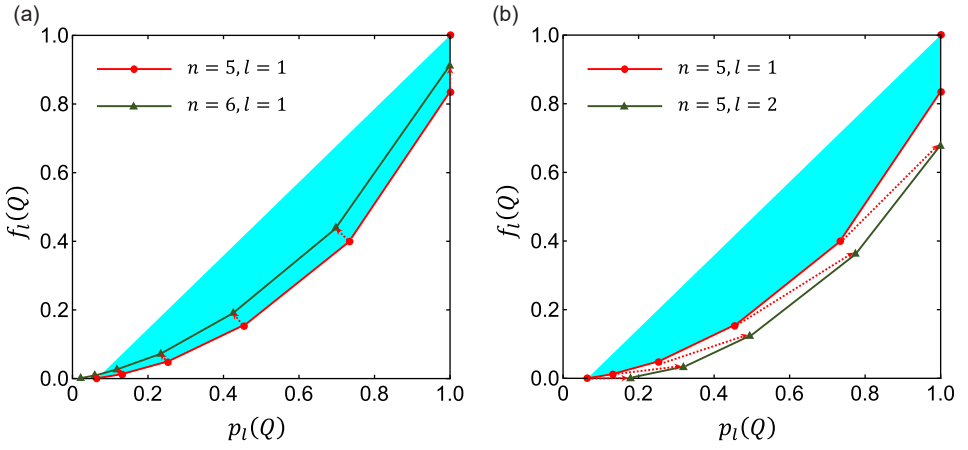


FIGURE K.1 (a) The lower boundary of $\mathcal{R}_{l,n,\lambda}$ is not higher than that of $\mathcal{R}_{l,n+1,\lambda}$ for $n \geq l+1$. The lower (upper) solid lines and circular (triangular) symbols represent the lower boundary and extremal points of $\mathcal{R}_{l=1,n=5,\lambda=1/e}$ ($\mathcal{R}_{l=1,n=6,\lambda=1/e}$), respectively. (b) The lower boundary of $\mathcal{R}_{l+1,n,\lambda}$ is not higher than that of $\mathcal{R}_{l,n,\lambda}$ for $n \geq l+2$. The upper (lower) solid lines and circular (triangular) symbols represent the lower boundary and extremal points of $\mathcal{R}_{l=1,n=5,\lambda=1/e}$ ($\mathcal{R}_{l=2,n=5,\lambda=1/e}$), respectively.

Combining (K.1) and (K.2), we have

$$\tau_{n,l}(h_z(l, n+1, \lambda)) \leq g_z(l, n+1, \lambda).$$

Therefore, the lower boundary of $\mathcal{R}_{n,l,\lambda}$ is not higher than that of $\mathcal{R}_{n+1,l,\lambda}$ for $n \geq l+1$. (see Figure K.1(a) for an illustration),

Monotonicity in l:

Step 1: To prove the monotonicity of $\overline{\epsilon}^{c}(T_l(\lambda), n, \delta)$ with l, we still consider two cases depending on the value of λ. First, when $\lambda = 0$, the monotonicity follows from (1). Second, when $\lambda > 0$, we show the monotonicity by showing that the lower boundary of $\mathcal{R}_{n,l+1,\lambda}$ is not higher than that of $\mathcal{R}_{n,l,\lambda}$ for $0 \leq l \leq n-2$. This fact can be shown as follows.

As shown later, we have the following relations:

$$g_{z_1}(l, n, \lambda) \leq g_{z_1}(l+1, n, \lambda), \quad h_{z_1}(l, n, \lambda) \leq h_{z_1}(l+1, n, \lambda), \tag{K.3}$$

$$h_{z+1}(l+1, n, \lambda) \geq h_z(l, n, \lambda), \tag{K.4}$$

$$\frac{g_{z+1}(l, n, \lambda) - g_z(l, n, \lambda)}{h_{z+1}(l, n, \lambda) - h_z(l, n, \lambda)} \geq \frac{g_{z+1}(l, n, \lambda) - g_{z+1}(l+1, n, \lambda)}{h_{z+1}(l, n, \lambda) - h_{z+1}(l+1, n, \lambda)} \tag{K.5}$$

for $z_1 = 0, 1, \ldots, n+1$, and $z = l, l+1, \ldots, n$.

Step 2: In the following, using (K.3)–(K.5), we show the following inequality by induction,

$$\tau_{n,l+1}(h_z(l, n, \lambda)) \leq g_z(l, n, \lambda). \tag{K.6}$$

First, we show (K.6) with $z = n+1$. Since (K.4) guarantees that $h_{n+1}(l, n, \lambda) \leq h_{n+1}(l+1, n, \lambda)$, we have

$$\tau_{n,l+1}(h_{n+1}(l, n, \lambda)) = 0 \overset{(a)}{=} g_{n+1}(l, n, \lambda),$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

where (a) follows from the definition of $g_z(l, n, \lambda)$. Hence, (K.6) holds when $z = n + 1$.

Next, we show (K.6) with $z = z_0$ by assuming that (K.6) holds when $z = z_0 + 1$. Equations (K.3)–(K.5) guarantee that

$$\kappa[A_{z_0+1}(l, n), A_{z_0+1}(l+1, n)](h_{z_0}(l, n, \lambda)) \le g_{z_0}(l, n, \lambda), \tag{K.7}$$

$$h_{z_0}(l, n, \lambda) \in [h_{z_0+1}(l, n, \lambda), h_{z_0+1}(l+1, n, \lambda)]. \tag{K.8}$$

Hence, we have

$$\begin{aligned}
&\tau_{n,l+1}(h_{z_0}(l, n, \lambda)) \\
&\overset{(a)}{\le} \kappa[(h_{z_0+1}(l, n, \lambda), \tau_{n,l+1}(h_{z_0+1}(l, n, \lambda))), A_{z_0+1}(l+1, n)](h_{z_0}(l, n, \lambda)) \\
&\overset{(b)}{\le} \kappa[A_{z_0+1}(l, n), A_{z_0+1}(l+1, n)](h_{z_0}(l, n, \lambda)),
\end{aligned} \tag{K.9}$$

where (a) follows from (K.8), and (b) follows from the assumption that (K.6) holds when $z = z_0 + 1$. The combination of (K.9) and (K.7) implies (K.6) with $z = z_0$.

Hence, we obtain (K.6) for $z = 1, \ldots, n + 1$, i.e. the lower boundary of $\mathcal{R}_{n,l+1,\lambda}$ is not higher than that of $\mathcal{R}_{n,l,\lambda}$ for $0 \le l \le n - 2$. (see Figure K.1(b) for an illustration).

Step 3: In the following, we shall prove (K.3), (K.4), and (K.5) in three steps, respectively.

Step 3.1: Proof of (K.3). Equation (K.3) follows from (76), (77), and Lemma 7.

Step 3.2: Proof of (K.4). For $z = l$, we have $h_{z+1}(l+1, n, \lambda) = h_z(l, n, \lambda) = 1$ according to (76). For $z = l + 1, \ldots, n$, we have

$$\begin{aligned}
(n+1)[h_{z+1}(l+1, n, \lambda) - h_z(l, n, \lambda)] &= [(n-z)B_{z+1,l+1} + (z+1)B_{z,l+1}] \\
&\quad - [(n-z+1)B_{z,l} + zB_{z-1,l}] \\
&= (n-z)(B_{z+1,l+1} - B_{z,l}) + z(B_{z,l+1} - B_{z-1,l}) \\
&\quad + (B_{z,l+1} - B_{z,l}) \\
&\overset{(a)}{=} (n-z)[(1-\lambda)B_{z,l} + \lambda B_{z,l+1} - B_{z,l}] \\
&\quad + z[(1-\lambda)B_{z-1,l} + \lambda B_{z-1,l+1} - B_{z-1,l}] \\
&\quad + (B_{z,l+1} - B_{z,l}) \\
&= (n-z)\lambda(B_{z,l+1} - B_{z,l}) + z\lambda(B_{z-1,l+1} - B_{z-1,l}) \\
&\quad + (B_{z,l+1} - B_{z,l}) \overset{(b)}{\ge} 0.
\end{aligned}$$

Here (a) follows from the basic relation $B_{z+1,l} = (1-\lambda)B_{z,l-1} + \lambda B_{z,l}$, and (b) follows from Lemma 7.

Step 3.3: Proof of (K.5). To prove (K.5), we shall consider three cases depending on the value of z. First, for $z = l$, we have $h_{z+1}(l+1, n, \lambda) = h_z(l, n, \lambda) = 1$ and $g_{z+1}(l+1, n, \lambda) \le g_z(l, n, \lambda)$, so (K.5) holds. Second, for $z = n$, the left-hand side of (K.5) is positive, and the right-hand side of (K.5) equals zero, so (K.5) holds.

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

Lastly, we consider the case $z \in \{l+1, \ldots, n-1\}$. To prove (K.5) in this case, it suffices to show

$$\frac{h_{z+1}(l,n,\lambda) - h_z(l,n,\lambda)}{g_{z+1}(l,n,\lambda) - g_z(l,n,\lambda)} - 1 \le \frac{h_{z+1}(l,n,\lambda) - h_{z+1}(l+1,n,\lambda)}{g_{z+1}(l,n,\lambda) - g_{z+1}(l+1,n,\lambda)} - 1 \tag{K.10}$$

for $z = l, l+1, \ldots, n$. Note that

$$\begin{aligned} \text{LHS of (K.10)} &= \frac{z(B_{z-1,l} - B_{z,l}) - B_{z,l}}{(n-z)(B_{z,l} - B_{z+1,l}) + B_{z,l}} \le \frac{z(B_{z-1,l} - B_{z,l})}{(n-z)(B_{z,l} - B_{z+1,l})} \\ &= \frac{z\Delta_{z-1,l}}{(n-z)\Delta_{z,l}} = \frac{z-l}{(n-z)\lambda}, \end{aligned} \tag{K.11}$$

where the last equality holds because

$$\frac{\Delta_{z-1,l}}{\Delta_{z,l}} = \frac{\binom{z-1}{l}}{\binom{z}{l}\lambda} = \frac{z-l}{z\lambda}.$$

In addition,

$$\begin{aligned} \text{RHS of (K.10)} &= \frac{(z+1)(B_{z,l+1} - B_{z,l})}{(n-z)(B_{z+1,l+1} - B_{z+1,l})} = \frac{(z+1)b_{z,l+1}}{(n-z)b_{z+1,l+1}} \\ &= \frac{z+1}{n-z} \cdot \frac{z-l}{\lambda(z+1)} = \frac{z-l}{(n-z)\lambda}. \end{aligned} \tag{K.12}$$

Then (K.11) and (K.12) together confirm (K.10), and complete the proof of (K.5).

Appendix L. Proof of Lemma 6

Proof of Lemma 6. When $0 < \delta \le B_{n,l}$, we have $z^* \ge n$ and $\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta) = 0$. So (83)–(85) hold obviously. In the following, we consider the case $B_{n,l} < \delta < 1$.

Step 1: Proof of (83). When $B_{n,l} < \delta < 1$, we have

$$\begin{aligned} \overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta) &\le \frac{g_{z_*}(l,n,\lambda)}{h_{z_*}(l,n,\lambda)} = \frac{1}{1 + \frac{z_*}{n-z_*+1} \cdot \frac{B_{z_*-1,l}}{B_{z_*,l}}} \\ &\le \frac{1}{1 + \frac{z_*}{n-z_*+1} \cdot \frac{z_*-l}{\lambda z_*}} = \frac{\lambda(n-z_*+1)}{\lambda(n-z_*+1) + z_* - l}, \end{aligned}$$

where the first inequality follows from Corollary 4 and the fact that $\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta) > 0$ when $B_{n,l} < \delta < 1$, and the last inequality follows from Lemma 10.

Step 2: Proof of (84). When $B_{n,l} < \delta < 1$, we have

$$\begin{aligned} \overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta) &\ge \frac{g_{z^*+1}(l,n,\lambda)}{h_{z^*+1}(l,n,\lambda)} = \frac{1}{1 + \frac{z^*+1}{n-z^*} \cdot \frac{B_{z^*,l}}{B_{z^*+1,l}}} \\ &\ge \frac{1}{1 + \frac{z^*+1}{n-z^*} \cdot \frac{1}{\lambda}} = \frac{\lambda(n-z^*)}{\lambda(n-z^*) + z^* + 1}, \end{aligned}$$

where the first inequality follows from Corollary 4; and the last inequality follows from Lemma 10.

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

Step 3: Proof of (85). When $B_{n,l} < \delta \le \frac{1}{2}$, we have $l/\nu \le z^* \le n$ according to Corollary 6, and

$$\begin{aligned}\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta) &\overset{(a)}{\ge} \frac{g_{z^*+1}(l, n, \lambda)}{h_{z^*+1}(l, n, \lambda)} = \frac{1}{1 + \frac{z^*+1}{n-z^*} \cdot \frac{B_{z^*,l}}{B_{z^*+1,l}}} \overset{(b)}{\ge} \frac{1}{1 + \frac{z^*+1}{n-z^*} \cdot \frac{z^*-l+1+\sqrt{\lambda l}}{(z^*+1)\lambda}} \\ &= \frac{\lambda(n-z^*)}{\lambda(n-z^*) + z^* - l + 1 + \sqrt{\lambda l}} = 1 - \frac{z^* - l + 1 + \sqrt{\lambda l}}{\lambda n + (\nu z^* - l + 1 + \sqrt{\lambda l})} \\ &\overset{(c)}{\ge} 1 - \frac{z^* - l + 1 + \sqrt{\lambda l}}{\lambda n},\end{aligned}$$

where (a) follows from Corollary 4; (b) follows from Lemma 10 and the inequality $z^* \ge l/\nu$; and (c) holds because $(\nu z^* - l + 1 + \sqrt{\lambda l}) \ge 0$. □

Appendix M. Proof of (88)

In the limit $n \to \infty$, we have

$$\begin{aligned}\frac{z_*}{n - z_* + 1} &\overset{(a)}{=} \frac{\nu s n - \Phi^{-1}(\delta)\sqrt{\lambda \nu s n} + O(1)}{\nu n - \nu s n + \Phi^{-1}(\delta)\sqrt{\lambda \nu s n} + O(1)} \\ &= \frac{\nu s - \Phi^{-1}(\delta)\sqrt{\lambda \nu s}\frac{1}{\sqrt{n}} + O\left(\frac{1}{n}\right)}{\nu - \nu s + \Phi^{-1}(\delta)\sqrt{\lambda \nu s}\frac{1}{\sqrt{n}} + O\left(\frac{1}{n}\right)} \\ &= \left[\nu s - \frac{\Phi^{-1}(\delta)\sqrt{\lambda \nu s}}{\sqrt{n}} + O\left(\frac{1}{n}\right)\right]\left[\frac{1}{\nu - \nu s} - \frac{\Phi^{-1}(\delta)\sqrt{\lambda \nu s}}{(\nu - \nu s)^2\sqrt{n}} + O\left(\frac{1}{n}\right)\right] \\ &= \frac{s}{1-s} - \frac{\Phi^{-1}(\delta)\sqrt{\lambda s}}{(1-s)^2\sqrt{\nu n}} + O\left(\frac{1}{n}\right), \qquad \text{(M.1)}\end{aligned}$$

where (a) follows from (86). Then, we have

$$\begin{aligned}\text{RHS of (87)} &= \left(1 + \frac{z_* B_{z_*-1, \lceil \nu s n \rceil}}{(n - z_* + 1) B_{z_*, \lceil \nu s n \rceil}}\right)^{-1} \\ &\overset{(a)}{=} \left(1 + \left[\frac{s}{1-s} - \frac{\Phi^{-1}(\delta)\sqrt{\lambda s}}{(1-s)^2\sqrt{\nu n}} + O\left(\frac{1}{n}\right)\right]\left[1 + \frac{\phi(\Phi^{-1}(\delta))\nu}{\delta\sqrt{\lambda \nu s n}} + O\left(\frac{1}{n}\right)\right]\right)^{-1} \\ &= \left(\frac{1}{1-s} + \frac{s}{1-s}\frac{\phi(\Phi^{-1}(\delta))\nu}{\delta\sqrt{\lambda \nu s n}} - \frac{\Phi^{-1}(\delta)\sqrt{\lambda s}}{(1-s)^2\sqrt{\nu n}} + O(n^{-1})\right)^{-1} \\ &= (1-s)\left(1 - \frac{s\phi(\Phi^{-1}(\delta))\nu}{\delta\sqrt{\lambda \nu s n}} + \frac{\Phi^{-1}(\delta)\sqrt{\lambda s}}{(1-s)\sqrt{\nu n}} + O(n^{-1})\right) \\ &= 1 - s + \frac{\Phi^{-1}(\delta)\sqrt{\lambda s}}{\sqrt{\nu n}} - \frac{s(1-s)\phi(\Phi^{-1}(\delta))\nu}{\delta\sqrt{\lambda \nu s n}} + O(n^{-1}) \\ &= 1 - s - C(\lambda, s, \delta)\sqrt{\frac{1}{n}} + O\left(\frac{1}{n}\right),\end{aligned}$$

which confirms (88). Here (a) follows from (B.6) in Lemma 18 and (M.1).

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

Appendix N. Proof of Theorem 2

We shall first prove (63) and then prove (65).

Proof of (63). In the following, we let $l = \lceil \nu k_0 \rceil$. The proof is composed of two steps.

Step 1: The aim of this step is to prove (63) in the case $B_{z^*,l} < \delta$.
In this case, we have

$$\lim_{n\to\infty} n\overline{\epsilon}(T_l(\lambda), n, \delta) = \lim_{n\to\infty} n[1 - \overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta)] \overset{(a)}{=} G(l, \delta, \lambda), \tag{N.1}$$

where (a) is shown in Section O by using Theorem 3. Then (N.1) confirms (63).

Step 2: The aim of this step is to prove (63) in the case $B_{z^*,l} = \delta$.
In this case, when $n \to \infty$, we have $B_{n,k} < \delta$ and $\hat{z} = z^*$. Then by a similar reasoning that leads to (N.1) we can deduce that

$$\lim_{n\to\infty} n\overline{\epsilon}(T_l(\lambda), n, \delta) = \frac{(B_{z_*,l} - \delta)z^* B_{z_*,l}}{\delta(B_{z_*,l} - B_{z^*,l})} = G(l, \delta, \lambda),$$

which confirms (63) again. □

Proof of (65). In the following, we let $l = \lceil \nu k_0 \rceil$, $\delta = \mathrm{e}^{-rn}$. The proof is composed of two steps.

Step 1: The aim of this step is to prove (65) in the case $r > \ln \lambda^{-1}$.
When $\delta = \mathrm{e}^{-rn}$ with $r > \ln \lambda^{-1}$, we have

$$\begin{aligned}\lim_{n\to\infty} \frac{\ln \delta}{\ln B_{n,l}} &\overset{(a)}{\geq} \lim_{n\to\infty} \frac{\ln(\mathrm{e}^{-rn})}{\ln\left[\frac{1}{\mathrm{e}\sqrt{l}}\mathrm{e}^{-D(l/n\|\nu)n}\right]} \\ &= \lim_{n\to\infty} \frac{r}{D(l/n\|\nu) + \frac{\ln(\mathrm{e}\sqrt{l})}{n}} \overset{(b)}{=} \frac{r}{\ln \lambda^{-1}} > 1,\end{aligned} \tag{N.2}$$

where (a) follows from the reverse Chernoff bound given in Lemma 8, and (b) holds because $D(0\|\nu) = \ln \lambda^{-1}$. (N.2) implies that $B_{n,l} > \delta$ when $n \to \infty$. This fact together with Theorem 3 implies that $\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta = \mathrm{e}^{-rn}) = 0$ as $n \to \infty$, which confirms (65).

Step 2: The aim of this step is to prove (65) in the case $0 < r \leq \ln \lambda^{-1}$.
When $\delta = \mathrm{e}^{-rn}$ with $0 < r \leq \ln \lambda^{-1}$, we have

$$\begin{aligned}\frac{\lambda(n - z^*)}{\lambda(n - z^*) + z^* - l + 1 + \sqrt{\lambda l}} &\leq \overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta) \\ &\leq \max\left\{0, \frac{\lambda(n - z_* + 1)}{\lambda(n - z_* + 1) + z_* - l}\right\}\end{aligned} \tag{N.3}$$

according to Lemma 6. In the following we shall show (65) by proving that

$$\lim_{n\to\infty} \text{LHS of (N.3)} = \lim_{n\to\infty} \text{RHS of (N.3)} = \frac{\lambda \ln \lambda + \lambda r}{\lambda \ln \lambda - \nu r}. \tag{N.4}$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

Note that

$$\lim_{n\to\infty} \text{LHS of} = \lim_{\delta\to 0} \frac{\lambda\left(\frac{n}{\ln\delta} - \frac{z^*}{\ln\delta}\right)}{\lambda\left(\frac{n}{\ln\delta} - \frac{z^*}{\ln\delta}\right) + \frac{z^*}{\ln\delta} + \frac{1-l+\sqrt{\lambda l}}{\ln\delta}} = \frac{\lambda\left(-\frac{1}{r} - \frac{1}{\ln\lambda}\right)}{\lambda\left(-\frac{1}{r} - \frac{1}{\ln\lambda}\right) + \frac{1}{\ln\lambda}} = \frac{\lambda\ln\lambda + \lambda r}{\lambda\ln\lambda - \nu r},$$

where the second equality follows from Lemma 16 and the relation $\delta = \mathrm{e}^{-rn}$. Using a similar reasoning, we can deduce that

$$\lim_{n\to\infty} \frac{\lambda(n - z_* + 1)}{\lambda(n - z_* + 1) + z_* - l} = \frac{\lambda\ln\lambda + \lambda r}{\lambda\ln\lambda - \nu r},$$

which is positive when $0 < r < \ln\lambda^{-1}$, and is equal to 0 when $r = \ln\lambda^{-1}$. These observations confirm (N.4) and complete the proof of (65). □

Appendix O. Proof of (N.1)

In the case $B_{z^*,l} < \delta$, when n goes to infinity, we have $B_{n,l} < \delta$ and

$$h_z(l, n, \lambda) \to B_{z,l} \tag{O.1}$$

for any finite $z \in \mathbb{Z}^{\geq l+1}$. Thus, in this limit we have $\hat{z} = z_*$, where $\hat{z}$ denotes the largest integer z such that $h_z(l, n, \lambda) \geq \delta$. This fact together with Theorem 3 implies that

$$\begin{aligned}
\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta) &\overset{(a)}{=} \frac{\tilde{\zeta}_\lambda(l, n, \delta, z_*)}{\delta} \\
&\overset{(b)}{=} \frac{1}{\delta}\left[(1 - \kappa_{z_*}(l, n, \delta, \lambda))g_{z^*}(l, n, \lambda) + \kappa_{z_*}(l, n, \delta, \lambda)g_{z_*}(l, n, \lambda)\right] \\
&\overset{(c)}{=} \frac{1}{\delta}\left(\frac{\left[\delta - h_{z^*}(l, n, \lambda)\right]g_{z_*}(l, n, \lambda)}{h_{z_*}(l, n, \lambda) - h_{z^*}(l, n, \lambda)} + \frac{\left[h_{z_*}(l, n, \lambda) - \delta\right]g_{z^*}(l, n, \lambda)}{h_{z_*}(l, n, \lambda) - h_{z^*}(l, n, \lambda)}\right) \\
&\overset{(d)}{=} \frac{(n+1)[\delta - h_{z^*}(l, n, \lambda)]\left[h_{z_*}(l, n, \lambda) - \frac{z_* B_{z_*-1,l}}{n+1}\right]}{\delta(n+1)[h_{z_*}(l, n, \lambda) - h_{z^*}(l, n, \lambda)]} \\
&\quad + \frac{(n+1)[h_{z_*}(l, n, \lambda) - \delta]\left[h_{z^*}(l, n, \lambda) - \frac{z^* B_{z_*,l}}{n+1}\right]}{\delta(n+1)[h_{z_*}(l, n, \lambda) - h_{z^*}(l, n, \lambda)]} \\
&= 1 - \frac{\delta z_* B_{z_*-1,l} - \delta z^* B_{z_*,l} + z^* B_{z_*,l} h_{z_*}(l, n, \lambda) - z_* B_{z_*-1,l} h_{z^*}(l, n, \lambda)}{\delta(n+1)[h_{z_*}(l, n, \lambda) - h_{z^*}(l, n, \lambda)]}.
\end{aligned}$$

Here (a) follows from Theorem 3, (b) follows from (80), (c) follows from (79), and (d) holds because

$$g_{z_*}(l, n, \lambda) = h_{z_*}(l, n, \lambda) - \frac{z_* B_{z_*-1,l}}{n+1}, \qquad g_{z^*}(l, n, \lambda) = h_{z^*}(l, n, \lambda) - \frac{z^* B_{z_*,l}}{n+1}.$$

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

It follows that

$$\begin{aligned}\lim_{n\to\infty} n\overline{\epsilon}(T_l(\lambda), n, \delta) &= \lim_{n\to\infty} n[1-\overline{\epsilon}^{\mathrm{c}}(T_l(\lambda), n, \delta)] \\ &= \lim_{n\to\infty} \frac{\delta z_* B_{z_*-1,l} - \delta z^* B_{z_*,l} + z^* B_{z_*,l} h_{z_*}(l,n,\lambda) - z_* B_{z_*-1,l} h_{z^*}(l,n,\lambda)}{\delta[h_{z_*}(l,n,\lambda) - h_{z^*}(l,n,\lambda)]} \\ &\overset{(a)}{=} \lim_{n\to\infty} \frac{\delta z_* B_{z_*-1,l} - \delta z^* B_{z_*,l} + z^* B_{z_*,l} B_{z_*,l} - z_* B_{z_*-1,l} B_{z^*,l}}{\delta(B_{z_*,l} - B_{z^*,l})} \\ &= \frac{(\delta - B_{z^*,l}) z_* B_{z_*-1,l} + (B_{z_*,l} - \delta) z^* B_{z_*,l}}{\delta(B_{z_*,l} - B_{z^*,l})} = G(l,\delta,\lambda),\end{aligned}$$

where (a) follows from (O.1).

Funding Information

The work at Fudan is supported by the National Natural Science Foundation of China (Grants No. 92165109 and No. 11875110) and Shanghai Municipal Science and Technology Major Project (Grant No. 2019SHZDZX01). MH is supported in part by the National Natural Science Foundation of China (Grants No. 62171212 and No. 11875110).

Competing Interests

There were no competing interests to declare which arose during the preparation or public cation process of this article.

References

[1] Ai, J., Kuželka, O. and Wang, Y. (2023). Hoeffding–Serfling inequality for U-statistics without replacement. *J. Theoret. Probab.* **36**, 390–408.
[2] Bardenet, R. and Maillard, O.-A. (2015). Concentration inequalities for sampling without replacement. *Bernoulli* **21**, 1361–1385.
[3] Bartroff, J., Lorden, G. and Wang, L. (2023). Optimal and fast confidence intervals for hypergeometric successes. *Am. Stat.* **77**, 151–159.
[4] Becchetti, L., Colesanti, U. M., Marchetti-Spaccamela, A. and Vitaletti, A. (2011). Recommending items in pervasive scenarios: Models and experimental analysis. *Knowl. Inf. Syst.* **28**, 555–578.
[5] Ben-Hamou, A., Peres, Y. and Salez, J. (2018). Weighted sampling without replacement. *Braz. J. Probab. Stat.* **32**, 657–669.
[6] Berry, A. C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Am. Math. Soc.* **49**, 122–136.
[7] Bradley, J. R. and Blossom, A. P. (2023). The generation of visually credible adversarial examples with genetic algorithms. *ACM Trans. Evol. Learn. Optim.* **3**, 1–44.
[8] Casiraghi, G. and Nanumyan, V. (2021). Configuration models as an urn problem. *Sci. Rep.* **11**, 13416.
[9] Chalé, M., Cox, B., Weir, J. and Bastian, N. D. (2023). Constrained optimization based adversarial example generation for transfer attacks in network intrusion detection systems. *Optim. Lett.* **18**, 2169–2188.
[10] Chiu, C.-H. (2023). A species richness estimator for sample-based incidence data sampled without replacement. *Methods Ecol. Evol.* **14**, 2189–2510.
[11] Choi, D. (2023). Estimating the prevalance of peer effects and other spillovers. arXiv: 2309.03969.
[12] Covey, R. and Buonamano, L. (2023). Survey design and estimating equations when combining big data with probability samples. arXiv: 2307.11999.
[13] Dembo, A. and Zeitouni, O. (2009). *Large Deviations Techniques and Applications*, Vol. **38**. Springer Science & Business Media.
[14] Dey, A. and Chaudhuri, P. (2023). A comparison of estimators of mean and its functions in finite populations. arXiv: 2305.15019.
[15] Dong, C., Liu, L. and Shang, J. (2022). Label noise in adversarial training: A novel perspective to study robust overfitting. *In Advances in Neural Information Processing Systems* **35**, pp. 17556–17567.

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

[16] FENG, D., DU, Y., GOMES, C. P. AND SELMAN, B. (2023). Weighted sampling without replacement for deep top-*k* classification. *Proceedings of the 40th International Conference on Machine Learning*, eds. A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato and J. Scarlett, Vol. 202 of *Proceedings of Machine Learning Research*. PMLR, pp. 9910–9920.
[17] GORDON, R. D. (1941). Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument. *Ann. Math. Stat.* **12**, 364–366.
[18] GREENE, E. AND WELLNER, J. A. (2016). Finite sampling inequalities: An application to two-sample Kolmogorov–Smirnov statistics. *Stochastic Process. Appl.* **126**, 3701–3715.
[19] GREENE, E. AND WELLNER, J. A. (2017). Exponential bounds for the hypergeometric distribution. *Bernoulli* **23**, 1911–1950.
[20] GUBRI, M., CORDY, M., PAPADAKIS, M., TRAON, Y. L. AND SEN, K. (2022). LGV: Boosting adversarial example transferability from large geometric vicinity. In *Computer Vision – ECCV 2022*, eds. S. Avidan, G. Brostow, M. Cissé, G. M. Farinella and T. Hassner. Springer Nature Switzerland, Cham. pp. 603–618.
[21] HALL, P. (1992). Principles of edgeworth expansion. In *The Bootstrap and Edgeworth Expansion*. Springer, pp. 39–81.
[22] HAYASHI, M. AND MORIMAE, T. (2015). Verifiable measurement-only blind quantum computing with stabilizer testing. *Phys. Rev. Lett.* **115**, 220502.
[23] HAYASHI, M. AND TSURUMARU, T. (2012). Concise and tight security analysis of the Bennett–Brassard 1984 protocol with finite key lengths. *New J. Phys.* **14**, 093014.
[24] HODARA, P. AND REYNAUD-BOURET, P. (2019). Exponential inequality for chaos based on sampling without replacement. *Stat. Probab. Lett.* **146**, 65–69.
[25] KAAS, R. AND BUHRMAN, J. M. (1980). Mean, median and mode in binomial distributions. *Stat. Neerl.* **34**, 13–18.
[26] KARAGIANNIDIS, G. K. AND LIOUMPAS, A. S. (2007). An improved approximation for the Gaussian Q-function. *IEEE Commun. Lett.* **11**, 644–646.
[27] LEHMANN, E. L. AND ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, Vol. 3. Springer Texts in Statistics.
[28] LEVINE, A. AND FEIZI, S. (2020). Robustness certificates for sparse adversarial attacks by randomized ablation. *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34, pp. 4585–4593.
[29] LI, Z., ZHU, H. AND HAYASHI, M. (2023). Robust and efficient verification of graph states in blind measurement-based quantum computation. *npj Quantum Inf.* **9**, 115.
[30] LODATO, M. A., WOODWORTH, M. B., LEE, S., EVRONY, G. D., MEHTA, B. K., KARGER, A., LEE, S., CHITTENDEN, T. W., D'GAMA, A. M., CAI, X., LUQUETTE, L. J., LEE, E., PARK, P. J. AND WALSH, C. A. (2015). Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94–98.
[31] MORIMAE, T., TAKEUCHI, Y. AND HAYASHI, M. (2017). Verification of hypergraph states. *Phys. Rev. A* **96**, 062321.
[32] MOTOYAMA, H. (2024). Extended Glivenko-Cantelli theorem for simple random sampling without replacement from a finite population. *Commun. Stat. Theory Methods* **53**, 5924–5934.
[33] NARODYTSKA, N. AND KASIVISWANATHAN, S. (2017). Simple black-box adversarial attacks on deep neural networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1310–1318.
[34] NOWAKOWSKI, S. (2021). Uniqueness of a median of a binomial distribution with rational probability. *Adv. Math. Sci. J.* **10**, 1951.
[35] O'NEILL, T. J. AND STERN, S. E. (2012). Finite population corrections for the Kolmogorov-Smirnov tests. *J. Nonparam. Stat.* **24**, 497–504.
[36] OUIMET, F. (2024). Deficiency bounds for the multivariate inverse hypergeometric distribution. *Stat. Papers* **65**, 3959–3969.
[37] PAPOULIS, A. (2002). *Probability, Random Variables, and Stochastic Processes*. *McGraw-Hill Europe*, New York, NY, USA.
[38] SAMBALE, H. AND SINULIS, A. (2022). Concentration inequalities on the multislice and for sampling without replacement. *J. Theoret. Probab.* **35**, 2712–2737.
[39] SAMOHYL, R. W. (2018). Acceptance sampling for attributes via hypothesis testing and the hypergeometric distribution. *J. Ind. Eng. Int.* **14**, 395–414.
[40] SERFLING, R. J. (1974). Probability inequalities for the sum in sampling without replacement. *Ann. Stat.* **2**, 39–48.
[41] SHAMIR, O. (2016). Without-replacement sampling for stochastic gradient methods. *In Advances in Neural Information Processing Systems* **29**, pp. 46–54.
[42] SIMON, H. U. AND TELLE, J. A. (2024). MAP-and MLE-based teaching. *J. Mach. Learn. Res.* **25**, 1–34.

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press

[43] SUTTER, T. M., MANDUCHI, L., RYSER, A. AND E. VOGT, J. (2022). Learning group importance using the differentiable hypergeometric distribution. arXiv: 2203.01629.
[44] TANASH, I. M. AND RIIHONEN, T. (2021). Improved coefficients for the Karagiannidis–Lioumpas approximations and bounds to the Gaussian Q-function. *IEEE Commun. Lett.* **25**, 1468–1471.
[45] WITHERS, C. S. AND NADARAJAH, S. (2023). Unbiased estimates for products of moments and cumulants for finite populations. *Mathematics* **11**, 3720.
[46] ZHU, H. AND HAYASHI, M. (2019). Efficient verification of pure quantum states in the adversarial scenario. *Phys. Rev. Lett.* **123**, 260504.
[47] ZHU, H. AND HAYASHI, M. (2019). General framework for verifying pure quantum states in the adversarial scenario. *Phys. Rev. A* **100**, 062335.
[48] ZHU, H., LI, Z. AND HAYASHI, M. (2022). Nearly tight universal bounds for the binomial tail probabilities. arXiv: 2211.01688.

https://doi.org/10.1017/apr.2025.10029 Published online by Cambridge University Press