Thompson Sampling for Adversarial Bit Prediction

Yuval Lewi YUVAL.LEWI@GMAIL.COM
Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv Israel

Haim Kaplan HAIMK@TAU.AC.IL
Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv Israel and Google Research, Tel Aviv

Yishay Mansour MANSOUR.YISHAY@GMAIL.COM
Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv Israel and Google Research, Tel Aviv

Editors: Aryeh Kontorovich and Gergely Neu

Abstract

We study the Thompson sampling algorithm in an adversarial setting, specifically, for adversarial bit prediction. We characterize the bit sequences with the smallest and largest expected regret. Among sequences of length T with $k < \frac{T}{2}$ zeros, the sequences of largest regret consist of alternating zeros and ones followed by the remaining ones, and the sequence of smallest regret consists of ones followed by zeros. We also bound the regret of those sequences, the worst case sequences have regret $O(\sqrt{T})$ and the best case sequence have regret $O(1)$.

We extend our results to a model where false positive and false negative errors have different weights. We characterize the sequences with largest expected regret in this generalized setting, and derive their regret bounds. We also show that there are sequences with $O(1)$ regret.

Keywords: Thompson sampling, Bit prediction, Adversarial setting, Regret, Multi-armed bandits

1. Introduction

Online learning and multi-arm bandits (MAB) are one of the most basic models for uncertainty, which are widely studied in machine learning. The main performance criteria used in this model is regret, which is the difference between the expected loss of the online algorithm, and the loss of the best algorithm from a benchmark class. (See, Cesa-Bianchi and Lugosi (2006); Bubeck and Cesa-Bianchi (2012); Lattimore and Szepesvári (2019); Slivkins (2019)). Bit prediction is one of the first problems for which online learning regret was analyzed (Cover, 1966), and has been extensively studied throughout the years (see, Rakhlin and Sridharan (2014)).

Thompson sampling (Thompson (1933)) is one of the earliest algorithms for MAB. It was originally motivated by a Bayesian setting, where the rewards are stochastic, and the reward of each action has a prior distribution. The algorithm maintains a posterior distribution for the reward of each action, and in each step, samples the posterior distribution of the mean reward of each action, and uses the action with the highest sampled value. In recent years, there has been a renewed interest in the Thompson sampling algorithm and its applications (see, Russo et al. (2018)), mainly due to its simplicity and good performance in practice.

Since Thompson sampling was designed for a Bayesian setting, it is natural to analyze its Bayesian regret (i.e., average the regret with respect to the prior). In many settings, we get an

elegant analysis and asymptotically optimal regret bounds. (See, Lattimore and Szepesvári (2019); Slivkins (2019); Russo and Roy (2016)).

While Thompson sampling was designed for a Bayesian setting, it was also recently analyzed in worst-case stochastic setting. More specifically, assume that the reward of each action is a Bernoulli random variable with unknown success probability. Unlike the Bayesian setting, there is no true prior over these parameters (success probabilities), and we want to bound the regret for the worst choice of the parameters. In this setting we start the Thompson sampling algorithm with a *fictitious* prior, say, a uniform distribution (of the success probability) for each action, and we update the posterior as though we were in the Bayesian setting. The works of Agrawal and Goyal (2017, 2013) show that Thompson sampling guarantees almost optimal regret bounds in the adversarial stochastic setting. Improved regret bounds which are parameter dependent are given in Kaufmann et al. (2012).

The papers mentioned above show the great success of Thompson sampling in stochastic settings, thus it is natural to investigate its performance in adversarial online model. In this model TS starts with a fictitious prior and an adversary selects the arbitrary input sequence. The completely adversatial model can be viewed as bounding the regret of the worst-case sequence possible, rather then the expected regret over some distribution in the stochastic settings. Specifically in this paper, our goal is to show that Thompson Sampling is successful for the adversarial bit sequence settings.

Our work considers the performance of Thompson sampling in an adversarial setting. Specifically, we consider the case of adversarial bit prediction, where the learner observes an arbitrary binary sequence, and at each time step predicts the next bit. The loss of the learner is the number of errors it makes, and the regret is the difference between the number of errors the online learner algorithm makes and best static bit prediction, i.e., the minimum between the number of ones and zeros in the sequence. We characterize the bit sequences on which Thompson sampling algorithm has the largest and smallest regret. We bound the regret of these sequences, and show that the worst case regret is $\Theta(\sqrt{T})$, for a sequence of length T, and best case regret of $\Theta(1)$.

More specifically, we initialize our Thompson sampling algorithm with a uniform (i.e., $\beta(1,1)$) prior distribution, and maintain a posterior beta distribution (whose parameters correspond to the number of ones and zeros seen so far). To predict the next bit, we draw a value from the beta posterior and predict one if the value is larger than $\frac{1}{2}$. Once we observe the bit we update our posterior.

For sequences of length T with $k \leq \frac{T}{2}$ zeros, we show that the sequences with the largest regret are of the form $\{01, 10\}^k 1^{T-2k}$, and the sequence with the smallest regret is $1^{T-k}0^k$ (for $k = \frac{T}{2}$ both sequences $1^{T/2}0^{T/2}$ and $0^{T/2}1^{T/2}$ have the same smallest regret). For example, if $k = 2$ and $T = 7$, the sequences with the largest regret are $0101111, 0110111, 1001111$ and 1010111, and the sequence with the smallest regret is 1111100. For $k > \frac{T}{2}$, we have the same characterization with 1 and 0 interchanged. We also bound the regret of these sequences and show that the expected regret on the worst case sequences is $\Theta(\sqrt{T})$ and that the expected regret on the best case sequences is $\Theta(1)$.

We extend the model to have different losses for false positive and false negative errors. Specifically, we have a trade-off parameter $q \in [0, 1]$ and we define the cost of a false positive to be q and the cost of a false negative to be $1 - q$. We call this extended model the *generalized bit-prediction* model. Note that for $q = \frac{1}{2}$ this loss is simply the number of errors multiplied by $\frac{1}{2}$, so this is a strict generalization our previous loss. Thompson sampling adapts naturally to the parameter q, by simply predicting one when the sampled value is larger than q (rather than larger than $\frac{1}{2}$). We

characterize for each $q \in [0,1]$ the bit sequences with the largest regret for this model and bound their regret. For example, for sequences of length $T = 100$ with 20 zeros and $q = \frac{1}{3}$, the worst case sequences are of the form $\{010, 001\}^{10}1^{70}$. In general, we show a family of bit-sequences with the highest regret for every trade-off parameter $q \in [0,1]$, number of zeros and number of ones. From that we conclude that the regret of Thompson sampling in the adversarial bit-prediction model is bounded by $O(\sqrt{q(1-q)T})$. We also show that there are sequences with regret equals or less then 1 without characterizing the best sequences.

Our work shows the great versatility of Thompson sampling. Namely, the same algorithm, with a prior of $\beta(1,1)$, can be analysed in Bayesian setting, when it is given the true prior, in an adversarial stochastic setting, when it is given a fictitious prior, and in the adversarial bit prediction problem, which we analyse in this work. Thompson sampling is not the only algorithm that achieves good performance both for adversarial and stochastic rewards (See, Bubeck and Slivkins (2012); Seldin and Slivkins (2014); Mourtada and Gaiffas (2018)), but it achieves this in a simple natural way, and as a side-product of a general Bayesian methodology, without trying to identify the nature of the environment.

1.1. Other related work

Adversarial bit prediction has a long history, starting with Cover (1966), and followed up by many additional works (see, Cesa-Bianchi and Lugosi (2006)). The exact min-max optimal strategy can be derived, when we view the problem as a zero-sum game (see, Rakhlin and Sridharan (2014)). The min-max optimal regret bound for the case of two actions was derived by Cover (1966) and for three actions by Gravin et al. (2016). Prediction of the next character in non-binary sequences has also received considerable attention, with respect to various benchmarks Feder et al. (1992); Cesa-Bianchi and Lugosi (1999). For the stochastic case, prediction of the next character in non-binary sequences was studied using Bayesian methods by Hutter (2003). Prediction of binary sequences with the log-loss in online adversarial environment has been studied by many due to its relation to data compression and information-theory (see for example, Freund et al. (1996), Merhav and Feder (1998) and Xie and Barron (2000)).

Adversarial online learning and multi-arm bandits have received significant attention in machine learning in the last two decades. (See the following books and surveys, Cesa-Bianchi and Lugosi (2006); Bubeck and Cesa-Bianchi (2012); Lattimore and Szepesvári (2019); Slivkins (2019)). A lower bound for the adversarial MAB problem was presented by Seldin and Lugosi. Notable results in adversarial online learning are the algorithm EXP3 (see, Auer et al. (2002b)) for adversarial bandits, the algorithm UCB1 (see, Auer et al. (2002a)) for stochastic bandits, and the regret analysis of the min-max algorithm (see, Audibert and Bubeck (2009)).

Thompson sampling has been studied in different environments over the years. In Gopalan (2013) it was observed that Thompson sampling with a Gaussian prior is equivalent to "Follow the Perturbed Leader" (FPL) of Kalai and Vempala (2005), and that fact was used to deduced the worst case regret of Thompson sampling with Gaussian distributions. A prior-dependent analysis was introduced by Russo and Roy (2016) using an information-theoretic tools, and the idea was expanded for first and second-order regret bounds by Bubeck and Sellke (2019).

Thompson sampling also showed good experimental results (see, Scott (2010); Chapelle and Li (2011)). Because of that, the algorithm is used in practice, with recommendation systems as an example (see, Kawale et al. (2015)). In Reinforcement Learning, a version of Thompson sampling

called "Posterior Sampling for Reinforcement Learning" (PSRL) is used (see, Osband et al. (2013); Osband and Van Roy (2017)). Bounds for the algorithm were proved in Agrawal and Jia (2017).

2. Model

A *bit prediction* game proceeds as follows. At time $t \in [T] = \{1, ..., T\}$ the learner outputs a bit $\hat{\gamma}_t \in \{0, 1\}$. Then, the learner observes a bit $\gamma_t \in \{0, 1\}$ and suffers a loss of $\ell(\hat{\gamma}_t, \gamma_t) = \mathbb{I}\{\hat{\gamma}_t \neq \gamma_t\}$.

We compare the loss of the online algorithm to a benchmark, which is the loss of the best static bit prediction. Given a bit sequence $\Gamma = (\gamma_1, ..., \gamma_T)$, let the number of ones up to t be $O_t(\Gamma) = |\{i \in [t] : \gamma_i = 1\}| = \sum_{i=1}^{t} \gamma_i$ and the number of zeros be $Z_t(\Gamma) = |\{i \in [t] : \gamma_i = 0\}| = \sum_{i=1}^{t} (1 - \gamma_i)$. The loss of the best static bit prediction is

$$static(\Gamma) = \min\left\{\sum_{t=1}^{T} \ell(1, \gamma_t), \sum_{t=1}^{T} \ell(0, \gamma_t)\right\} = \min\{Z_T(\Gamma), O_T(\Gamma)\}.$$

The goal of the learner is to minimize the regret, which is the difference between the online cumulative loss and the loss of the best static bit prediction. Specifically, for an algorithm A,

$$Regret_A(\Gamma) = \sum_{t=1}^{T} E_{\hat{\gamma}_t \sim A}[\ell(\hat{\gamma}_t, \gamma_t) \mid \Gamma] - static(\Gamma),$$

where $\Gamma \in \{0, 1\}^T$ is a fixed bit sequence, and the expectation is taken over the predictions of algorithm A. We extend the standard bit prediction game and define a *generalized bit prediction* game, where the false positive (FP) and false negative (FN) errors have different weights.[1] Given a *trade-off parameter* $q \in [0, 1]$, we define a loss ℓ^q, as follows,

$$\ell^q(\hat{\gamma}_t, \gamma_t) = q\mathbb{I}\{\hat{\gamma}_t = 1, \gamma_t = 0\} + (1 - q)\mathbb{I}\{\hat{\gamma}_t = 0, \gamma_t = 1\}.$$

Namely, the false positive errors are weighted by q while the false negative errors are weighted by $1 - q$. Note that for $q = \frac{1}{2}$, for any $(\hat{\gamma}_t, \gamma_t)$ we have that $\ell^{1/2}(\hat{\gamma}_t, \gamma_t) = \frac{1}{2}\ell(\hat{\gamma}_t, \gamma_t)$, so for $q = \frac{1}{2}$ the extended loss is essentially the 0-1 loss.

Similarly, the benchmark for the generalized bit prediction is the best static bit prediction, namely,

$$static^q(\Gamma) = \min\left\{\sum_{t=1}^{T} \ell^q(1, \gamma_t), \sum_{t=1}^{T} \ell^q(0, \gamma_t)\right\} = \min\{qZ_T(\Gamma), (1 - q)O_T(\Gamma)\},$$

and the regret of algorithm A on a given bit sequence $\Gamma \in \{0, 1\}^T$ is

$$Regret_A^q(\Gamma) = \sum_{t=1}^{T} E_{\hat{\gamma}_t \sim A}\left[\ell^q(\hat{\gamma}_t, \gamma_t) \mid \Gamma\right] - static^q(\Gamma).$$

1. A false positive error is when the learner predicts $\hat{\gamma}_t = 1$ and $\gamma_t = 0$, and false negative error is when $\hat{\gamma}_t = 0$ and $\gamma_t = 1$.

2.1. Distributions

We use extensively the Beta distribution, denoted by $\beta(a,b)$, where $a,b > 0$, and the Binomial distribution, denoted by $Bin(n,p)$ where n is the number of trials and $p \in [0,1]$ is the success probability. We denote by $Ber(p)$ a Bernoulli random variable with success probability $p \in [0,1]$. For a distribution D, the Cumulative Distribution Function (CDF) is denoted by F_D.

The following identity is a well known fact related to the the Beta distribution (see, DLMF, Eq. 8.17.4)

Fact 1 *For $a,b \in \mathbb{N}^+$ and $p \in [0,1]$ we have $F_{\beta(a,b)}(p) = 1 - F_{\beta(b,a)}(1-p)$.*

The $\beta(a,b)$ distribution is widely used in Bayesian setting to define the uncertainty over the parameter p of a Bernoulli random variable $Ber(p)$. The distribution $\beta(1,1)$, which is the uniform distribution over $[0,1]$, is used as the prior distribution of p. Given $a+b$ observations of the random variable $Ber(p)$, where a is the number of realizations which are 1 and b is the number of realizations which are 0, then the posterior distribution of p is $\beta(a+1,b+1)$ (assuming the prior distribution is $\beta(1,1)$).

The following is a well known property of the CDF of the Beta distribution.

Fact 2 *(DLMF, Eq. 8.17.20-21) For every $x \in [0,1]$ and $a,b \in \mathbb{R}$ s.t. $a,b > 0$, the following holds*

$$F_{\beta(a+1,b)}(x) = F_{\beta(a,b)} - \frac{x^a(1-x)^b}{a\mathrm{B}(a,b)} \qquad \textit{and} \qquad F_{\beta(a,b+1)}(x) = F_{\beta(a,b)}(x) + \frac{x^a(1-x)^b}{b\mathrm{B}(a,b)}$$

where $B(a,b) = \frac{(a-1)!(b-1)!}{(a+b-1)!}$ is the Beta function.

For the analysis we use the following theorems regarding the tail of the $\beta(a,b)$ distribution, when we fix the parameter $b = n+1$ and sum over parameters $a \geq 1$.

Theorem 3 *For every $n \geq 1$ we have $\sum_{i=n+1}^{\infty} F_{\beta(i+1,n+1)}\left(\frac{1}{2}\right) = O(\sqrt{n})$.*

2.2. Notations

When the bit sequence $\Gamma = (\gamma_1, \ldots, \gamma_T)$ can be inferred from the context, we use O_t and Z_t rather than $O_t(\Gamma)$ and $Z_t(\Gamma)$.

We also define the $sign$ function as $sign(x) = \begin{cases} 1 & x>0 \\ 0 & x=0 \\ -1 & x<0 \end{cases}$.

For functions $f,g \in \mathbb{R} \to \mathbb{R}$ we denote $g = O(f)$ iff there exist $c_1, c_2 \in \mathbb{R}$ such that $g(x) \leq c_1 f(x) + c_2$ for every $x \in \mathbb{R}$.

3. Thompson sampling for bit prediction

The Thompson sampling algorithm requires a prior distribution for its initialization. Given the observations, it updates the prior distribution to a posterior distribution. The learner samples the posterior distribution, and thresholds the sampled value at half (for bit prediction) or q (for generalized bit prediction).

More specifically. We consider the prior distribution $\beta(1,1)$, which is a uniform distribution over $[0,1]$. Note that this prior is fictitious, and used only to initialize the Thompson sampling

Algorithm 1: Thompson sampling with Beta prior for bit prediction
input : Trade-off parameter $q \in [0,1]$.
initialize: Set $O_0 = 0, Z_0 = 0$.

for *each time* t *in* $[T]$ **do**
  Sample x_t *from the* $\beta\left(O_{t-1}+1, Z_{t-1}+1\right)$ *distribution.*
  Predict bit $\hat{\gamma}_t = \mathbb{I}\{x_t > q\}$.
  Observe bit γ_t *and suffer loss* $\ell_t = \ell^q(\hat{\gamma}_t, \gamma_t)$.
  Update $O_t = O_{t-1} + \gamma_t$ *and* $Z_t = Z_{t-1} + (1-\gamma_t)$.
end

algorithm. At time t the learner samples a value x_t from the distribution $\beta(O_{t-1}+1, Z_{t-1}+1)$, where O_{t-1} and Z_{t-1} are the number of observed 1's and 0's up to time $t-1$, respectively. At time t the learner predicts $\hat{\gamma}_t = \mathbb{I}\{x_t > q\}$, where q is the trade-off parameter of the loss. Then the learner observes the feedback bit γ_t and suffers loss $\ell^q(\hat{\gamma}_t, \gamma_t)$. The resulting Thompson sampling algorithm is described in Algorithm 1, and in the analysis we refer to this algorithm as $TS(q)$.

In Section 4 we prove the "Swapping Lemma", which analyses the effect of a single swap on the regret, which allows us to identify the sequences with the largest and smallest regret. In Section 5 we bound the regret of these sequences, thereby obtaining tight upper and lower bounds on the regret. Section 6 addresses the generalized bit prediction case.

4. Swapping lemma

In this section we compare the regret of two bit sequences which differ by a single swap. This is an essential building block in our analysis of the worst case and the best case regret of the Thompson sampling algorithm.

Swap operation: Given a bit sequence $\Gamma = (\gamma_1, \ldots, \gamma_T)$, performing the swap operation at position $t \in [T]$ results in a sequence that swaps γ_t and γ_{t+1} in Γ and keeps all other bits unchanged. Formally, $Swap(\Gamma, t) = (\gamma_1, \ldots, \gamma_{t-1}, \gamma_{t+1}, \gamma_t, \gamma_{t+2}, \ldots, \gamma_T)$.

The swapping lemma that compares the regret of Thompson sampling, $TS(q)$, on the bit sequences Γ and $Swap(\Gamma, t)$.

To illustrate the swapping lemma consider the case $q = \frac{1}{2}$, so $\frac{q}{1-q} = 1$. If we had more zeros up to position t-1 then having the one earlier increases the regret. If we had more ones up to position t-1 then having zero earlier increases the regret. More precisely, for each t such that $\gamma_t = 0$, $\gamma_{t+1} = 1$ and $O_{t-1} < Z_{t-1}$, swapping γ_t and γ_{t+1} increases the regret. Similarly, if $\gamma_t = 1$, $\gamma_{t+1} = 0$ and $O_{t-1} > Z_{t-1}$ then swapping γ_t and γ_{t+1} increases the regret. In other words,

Lemma 4 (Swapping Lemma) *Fix a bit sequence* $\Gamma = (\gamma_1, \ldots, \gamma_T) \in \{0,1\}^T$. *For every* t, *such that* $\gamma_t = 0$ *and* $\gamma_{t+1} = 1$, *we have*

$$Regret^q_{TS(q)}(\Gamma) < Regret^q_{TS(q)}(Swap(\Gamma, t)) \iff \frac{q}{1-q} > \frac{O_{t-1}+1}{Z_{t-1}+1}.$$

For every t, *such that* $\gamma_t = 1$ *and* $\gamma_{t+1} = 0$, *we have*

$$Regret^q_{TS(q)}(\Gamma) < Regret^q_{TS(q)}(Swap(\Gamma, t)) \iff \frac{q}{1-q} < \frac{O_{t-1}+1}{Z_{t-1}+1}.$$

In addition,

$$Regret^q_{TS(q)}(\Gamma) = Regret^q_{TS(q)}(Swap(\Gamma,t)) \iff \frac{q}{1-q} = \frac{O_{t-1}+1}{Z_{t-1}+1}.$$

Proof Sketch We consider the difference between the regret of $TS(q)$ on the bit sequence Γ and on the bit sequence $Swap(\Gamma,t)$. The two bit sequences differ only at locations t and $t+1$. Since the benchmark of a sequence depends only on the total number of zeros and ones in the sequence, the benchmarks on Γ and $Swap(\Gamma,t)$ are identical, i.e., $static^q(\Gamma) = static^q(Swap(\Gamma,t))$. Therefore, the difference between the regrets is equals to the difference between the losses at time t and $t+1$.

Consider time $t \in [T]$ such that $\gamma_t = 0$ and $\gamma_{t+1} = 1$.Using the insights above it is easy to show that,

$$\begin{aligned} &Regret^q_{TS(q)}(\Gamma) - Regret^q_{TS(q)}(Swap(\Gamma,t)) \\ &\quad = (1-q)F_{\beta(O_{t-1}+1,Z_{t-1}+2)}(q) + qF_{\beta(O_{t-1}+2,Z_{t-1}+1)}(q) - F_{\beta(O_{t-1}+1,Z_{t-1}+1)}(q), \end{aligned}$$

Using the recurrence relations in Fact 2 we show that,

$$\begin{aligned} &Regret^q_{TS(q)}(\Gamma) - Regret^q_{TS(q)}(Swap(\Gamma,t)) \\ &\quad = \frac{q^{O_{t-1}+1}(1-q)^{Z_{t-1}+1}}{B(O_{t-1}+1, Z_{t-1}+1)}\left(\frac{1-q}{Z_{t-1}+1} - \frac{q}{O_{t-1}+1}\right), \end{aligned}$$

Since $\frac{q^{O_{t-1}+1}(1-q)^{Z_{t-1}+1}}{B(O_{t-1}+1,Z_{t-1}+1)} > 0$, we have

$$Regret^q_{TS(q)}(\Gamma) < Regret^q_{TS(q)}(Swap(\Gamma,t)) \iff \frac{q}{1-q} > \frac{O_{t-1}+1}{Z_{t-1}+1},$$

and equality holds iff $\frac{q}{1-q} = \frac{O_{t-1}+1}{Z_{t-1}+1}$. The second case, where $\gamma_t = 1$ and $\gamma_{t+1} = 0$, is similar. ■

5. Regret characterization for $q = \frac{1}{2}$

In this section we use the swapping lemma to characterize the sequences on which $TS(\frac{1}{2})$ has the largest and smallest regret. We denote by k the number of zeros in the sequence and characterize the sequences of worst and best regret for each k. Notice that we may assume that $k \le \frac{T}{2}$ since any sequence Γ has the same regret as the sequence Γ' obtained from Γ by flipping each bit. Indeed, $static(\Gamma) = static(\Gamma')$ and the expected loss of $TS(\frac{1}{2})$ on Γ and Γ' is the same (by Fact 1).

5.1. Worst-case regret

Consider bit sequences $\Gamma = (\gamma_1, \ldots, \gamma_T)$ with k zeros, where $k \le \frac{T}{2}$. We first show that among these bit sequences the ones of largest regret are of the form $\{01,10\}^k 1^{T-2k}$. Then, we prove that the regret of each of these sequences is $\Theta(\sqrt{k})$.

Theorem 5 *For any $\Gamma_1, \Gamma_2 \in \{01,10\}^k 1^{T-2k}$ we have $Regret^{1/2}_{TS(\frac{1}{2})}(\Gamma_1) = Regret^{1/2}_{TS(\frac{1}{2})}(\Gamma_2)$. In addition, for any $\Gamma_3 \notin \{01,10\}^k 1^{T-2k}$ we have $Regret^{1/2}_{TS(\frac{1}{2})}(\Gamma_1) > Regret^{1/2}_{TS(\frac{1}{2})}(\Gamma_3)$.*

Proof Note that for any $i \in [k]$ we have $O_{2i}(\Gamma_1) = Z_{2i}(\Gamma_1) = i$. By Lemma 4 this implies that $Regret^{1/2}_{TS(\frac{1}{2})}(\Gamma_1) = Regret^{1/2}_{TS(\frac{1}{2})}(Swap(\Gamma_1, i))$. Since we can transform Γ_1 to Γ_2 by a sequence of swap operations at certain locations $2i$, it follows that $Regret^{1/2}_{TS(\frac{1}{2})}(\Gamma_1) = Regret^{1/2}_{TS(\frac{1}{2})}(\Gamma_2)$. This implies that all the sequences of the form $\{01, 10\}^k 1^{T-2k}$ have the same regret.

Let $\Gamma_3 = (\gamma_1, \ldots, \gamma_T) \in \{0, 1\}^T$ be a bit sequence of length T with k zeros such that $\Gamma_3 \notin \{01, 10\}^k 1^{T-2k}$. We show that for some $t \in [T]$, the sequence $Swap(\Gamma_3, t)$ has a regret larger than Γ_3.

Since $\Gamma_3 \notin \{01, 10\}^k 1^{T-2k}$, there is an index $i \leq k-1$ such that either $\gamma_{2i+1} = \gamma_{2i+2} = 1$ or $\gamma_{2i+1} = \gamma_{2i+2} = 0$. Let i to be the smallest such index. Assume that $\gamma_{2i+1} = \gamma_{2i+2} = 1$. (The case of $\gamma_{2i+1} = \gamma_{2i+2} = 0$ is similar.) It follows that $O_{2i} = Z_{2i}$ and $O_{2i+1} = Z_{2i+1} + 1$. Let $j > 2i + 2$ be the minimal index such that $\gamma_j = 0$. Such an index must exist, since there are k zeros in Γ_3 and until index $2i$ there were only $i \leq k-1$ zeros. Since $\gamma_{j-1} = \gamma_{j-2} = 1$ we have $\frac{O_{j-1}}{Z_{j-1}} > \frac{O_{j-2}}{Z_{j-2}} \geq \frac{O_{2i+1}}{Z_{2i+1}} > 1$. By Lemma 4, the sequence $Swap(\Gamma_3, j-1)$ has regret larger than Γ_3, i.e., $Regret^{1/2}_{TS(\frac{1}{2})}(\Gamma_3) < Regret^{1/2}_{TS(\frac{1}{2})}(Swap(\Gamma_3, t))$.

Since there are finite number of bit sequences of length T with k zeros, we get that sequences with the largest regret must be of the form $\{01, 10\}^k 1^{T-2k}$. ■

Given the above theorem, to bound the worst case regret of $TS(\frac{1}{2})$, we can focus on the sequence $W^k_T = \{01\}^k 1^{T-2k}$ and bound $Regret^{1/2}_{TS(\frac{1}{2})}(W^k_T)$.

Theorem 6 *For every $T \in \mathbb{N}^+$ and $k \leq \frac{T}{2}$ we have, $Regret^{1/2}_{TS(\frac{1}{2})}(W^k_T) = \Theta(\sqrt{k})$.*

Proof Sketch Let $W^k_T = (w_1, \ldots, w_T)$, where we have: (1) $w_t = 0$ for $t \in A_1 = \{2i-1 \mid i \in [k]\}$, (2) $w_t = 1$ for $t \in A_2 = \{2i \mid i \in [k]\}$, and (3) $w_t = 1$ for $t \in A_3 = \{i \mid i \geq 2k+1\}$. We bound the expected number of errors made by $TS(\frac{1}{2})$ on each of these three subsets. Then, from these bounds we derive a bound on the loss and the regret. Specifically we prove the following:

1. For $t \in A_1$, $Z_t = O_t$ and thus the probability to predict the next bit is $\frac{1}{2}$. Therefore, the expected number of false positive errors in A_1 is
$$\sum_{t=1}^{k} \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq w_t\} \mid W^k_T\right] = \frac{k}{2}.$$
2. For $t \in A_2$, $Z_t = O_t + 1$ and the difference between the probability to predict 0 and the probability to predict 1 is small and can be bounded. Therefore, the expected number of false negative errors in A_2 is
$$\sum_{i=1}^{k} \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_{2i} \neq w_{2i}\} \mid W^k_T\right] = \frac{k}{2} + \Theta(\sqrt{k}).$$
3. The expected number of false negative in A_3 is show to be
$$\sum_{t=2k+1}^{T} \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq w_t\} \mid W^k_T\right] = \sum_{t=2k+1}^{T} F_{\beta(t-k+1,k+1)}\left(\frac{1}{2}\right) = O(\sqrt{k}),$$

where the last equality follows from Theorem 3.

Summing up the errors over A_1, A_2, and A_3, and recalling that the static prediction makes $\min\{T-k,k\}=k$ errors, we bound the regret as follows

$$\sum_{t=1}^{T}\mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t\neq w_t\}\mid W_T^k\right]-\min\{T-k,k\}=\frac{k}{2}+\left(\frac{k}{2}+\Theta(\sqrt{k})\right)+O(\sqrt{k})-k=\Theta(\sqrt{k}).$$

■

Since $k\leq\frac{T}{2}$, we have the following corollary.

Corollary 7 *For any sequence of length T, the regret of $TS(\frac{1}{2})$ is at most $O(\sqrt{T})$.*

Remark 8 *Note that in fact we proved that $Regret^{1/2}_{TS(\frac{1}{2})}(\Gamma)=\Theta(\sqrt{\min\{O_T(\Gamma),Z_T(\Gamma)\}})$.*

5.2. Best-case regret

In this subsection, we characterize the sequences with the lowest regret and bound them.

Theorem 9 *The bit sequence with the lowest regret of length T with $k<\frac{T}{2}$ zeros is $B_T^k=1^{T-k}0^k$. For $k=\frac{T}{2}$, both $1^{T/2}0^{T/2}$ and $0^{T/2}1^{T/2}$ have the lowest regret.*

We now bound the regret of B_T^k.

Theorem 10 *For every $T\in\mathbb{N}^+$ and $k\leq\frac{T}{2}$ we have, $Regret^{1/2}_{TS(\frac{1}{2})}(B_T^k)\leq 1$, where $B_T^k=1^{T-k}0^k$.*

6. Regret characterization for a general q

To get some intuition regarding this generalization to an arbitrary trade-off parameter q consider the following simple example. Assume that $q=\frac{1}{3}$, and thereby $\frac{q}{1-q}=\frac{1}{2}$ and lets construct a sequence such that we cannot increase the regret by swapping any pair of consecutive bits. This sequence cannot start with a 1, since if it does then by the swapping lemma (Lemma 4 we will be able to increase the regret by swapping the first 0 with the 1 preceding it. So we must start with a 0. In general we determine bit $t+1$ by comparing $\frac{O_t+1}{Z_t+1}$ to $\frac{1}{2}$ (i.e., $\frac{q}{1-q}$). If they are equal then the bit in position $t+1$ is either 0 or 1. If $\frac{O_t+1}{Z_t+1}>\frac{1}{2}$ the bit in position $t+1$ is 0 since otherwise we will be able to increase the regret by swapping the first 0 following position $t+1$ with its preceding 1. Similarly, if $\frac{O_t+1}{Z_t+1}<\frac{1}{2}$ the bit in position $t+1$ is 1 since otherwise we will be able to increase the regret by swapping the first 1 following position $t+1$ with its preceding 0.

It follows that the second bit could be either 0 or 1 since $\frac{O_1+1}{Z_1+1}=\frac{q}{1-q}=\frac{1}{2}$. If we have a 0 at position 2 then $\frac{O_2+1}{Z_2+1}=\frac{1}{3}<\frac{1}{2}$ and therefore we must continue with a 1 at position 3. Then we have that $\frac{O_3+1}{Z_3+1}=\frac{2}{3}>\frac{1}{2}$ so we put 0 at position 4, and we are back in the situation where $\frac{O_4+1}{Z_4+1}=\frac{1}{2}$ so we can choose either 0 or 1 at position 5. Similarly, if we place a 1 at position 2 then we will have to continue with two 0's and then we will be free to choose at position 5 either 0 or 1. It follows

that the family of sequences of the form $0\{100,010\}^* x\{1^*,0^*\}$ (where x could be any prefix of 100 or 010) contains all sequences of largest regret. (We will in fact show that they all have the same regret.)

To gain some deeper intuition assume now that q is a rational number and $\frac{q}{1-q}=\frac{n_1}{n_2}$ (where n_1 and n_2 do not have common divisors) and lets try to construct a sequence that we cannot increase its regret by applying the swapping lemma. Whenever $\frac{O_t+1}{Z_t+1}=\frac{n_1}{n_2}$ we can choose any bit to position $t+1$. At this point we have that $n_2(O_t+1)=n_1(Z_t+1)$ and therefore $n_1(Z_t+1)$ is a multiple of n_2 and $n_2(O_t+1)$ is a multiple of n_1. Once we choose, say 0, then we are forced to choose a particular sequence in the following n_1+n_2-1 steps, until we will again have that $n_2(O_{t'}+1)=n_1(Z_{t'}+1)$ for $t'=t+n_1+n_2$ among these bits n_2 would be zeros and n_1 would be ones so $Z_{t'}=Z_t+n_2$ $O_{t'}=O_t+n_1$.

The structure of this section is similar to the structure of Section 5. First, we characterize the bit sequences of largest regret. Then, we bound the regret of these sequences.

6.1. Worst-case sequences

Consider the following function that maps a bit-sequence to a set of bits

$$\forall \Phi \in \{0,1\}^* : H^q(\Phi) = \begin{cases} \{0\} & \frac{O(\Phi)+1}{Z(\Phi)+1} > \frac{q}{1-q} \\ \{1\} & \frac{O(\Phi)+1}{Z(\Phi)+1} < \frac{q}{1-q} \\ \{0,1\} & \frac{O(\Phi)+1}{Z(\Phi)+1} = \frac{q}{1-q} \end{cases}, \tag{1}$$

where $O(\Phi)$ is the total number of 1s in Φ and $Z(\Phi)$ is the total number of 0s in Φ.

For every sequence $\Gamma=(\gamma_1,\ldots,\gamma_T)\in\{0,1\}^T$ we define $p(\Gamma)$ to be the largest index t s.t. $\forall i\in[t]:\gamma_i\in H^q(\Gamma_{1:i-1})$, where $\Gamma_{1:n}=(\gamma_1,\ldots,\gamma_n)$. We call a bit sequence $\Gamma=(\gamma_1,\ldots,\gamma_T)$ a *worst-case sequence* if $\gamma_{p(\Gamma)+1}=\ldots=\gamma_T$. We define the subsequence $(\gamma_1,\ldots,\gamma_{p(\Gamma)})$ as the *head* of Γ and denote it $head(\Gamma)$ and the subsequence $(\gamma_{p(\Gamma)+1},\ldots,\gamma_T)$ as the *tail* of Γ and denote it $tail(\Gamma)$.

For start, we characterize the tail of a worst-case sequence.

Theorem 11 *Let Γ be a worst-case sequence. If $Z_T\le(1-q)T-q$ then the $tail(\Gamma)$ is filled with ones. Otherwise, the $tail(\Gamma)$ is filled with zeros.*

6.2. Worst-case regret

In this subsection we prove that all the worst-case sequences have the largest regret and prove an upper bound on this regret.

Theorem 12 *Let $\Gamma\in\{0,1\}^T$, s.t. Γ is not a worst-case sequence. Then, there exists $t\in[T]$ such that $Regret^q_{TS(q)}(\Gamma)<Regret^q_{TS(q)}(Swap(\Gamma,t))$.*

Proof Let $i=p(\Gamma)+1$. Since Γ is not a worst-case sequence, there is an index $j>i$ such that $\gamma_j\neq\gamma_i$ (since, from Theorem 11, $tail(\Gamma)$ contains both 0's and 1's). Assume j is the smallest index with this property.

Case 1 Assume $\gamma_i=0$ and $\gamma_j=1$. Since $\gamma_i\notin H^q(\Gamma_{1:i-1})$ we have $\frac{O_{i-1}(\Gamma)+1}{Z_{i-1}(\Gamma)+1}<\frac{q}{1-q}$. From the definition of j follows that $\gamma_i=\gamma_{i+1}=\ldots=\gamma_{j-1}=0$ and thus $\frac{O_{j-2}(\Gamma)+1}{Z_{j-2}(\Gamma)+1}\le\frac{O_{i-1}(\Gamma)+1}{Z_{i-1}(\Gamma)+1}<\frac{q}{1-q}$. By Lemma 4, the sequence $Swap(\Gamma,j-1)$ has a regret larger than Γ.

Case 2 Assume $\gamma_i = 1$ and $\gamma_j = 0$. Since $\gamma_i \notin H^q(\Gamma_{1:i-1})$ we have $\frac{O_{i-1}(\Gamma)+1}{Z_{i-1}(\Gamma)+1} > \frac{q}{1-q}$. From the definition of j follows that $\gamma_i = \gamma_{i+1} = \ldots = \gamma_{j-1} = 1$ and thus $\frac{O_{j-2}(\Gamma)+1}{Z_{j-2}(\Gamma)+1} \geq \frac{O_{i-1}(\Gamma)+1}{Z_{i-1}(\Gamma)+1} > \frac{q}{1-q}$. By Lemma 4, the sequence $Swap(\Gamma, j-1)$ has a regret larger than Γ. ■

Theorem 12 implies that any sequence of largest regret is a worst-case sequence. Next we prove that all worst-case sequences of length T with k zeros have the same regret.

Lemma 13 *All the worst-case sequences of length T with k zeros have the same regret.*

Let $W_T^k = (w_1, \ldots, w_T) \in \{0,1\}^T$ be a worst-case sequence with k zeros such that for all $t \leq p(W_T^k)$ with $\frac{O_{t-1}+1}{Z_{t-1}+1} = \frac{q}{1-q}$ we have $\gamma_t = 0$. Since by Lemma 13 all the worst-case sequences with the same number of zeros have the same regret, we can focus on bounding the regret of W_T^k.

Theorem 14 *For every $T \in \mathbb{N}^+$, $q \in \left[0, \frac{1}{2}\right]$ and k zeros we have*

$$Regret^q_{TS(q)}(W_T^k) = \begin{cases} O(\sqrt{qk}) & k \leq (1-q)T - q \\ O(\sqrt{(1-q)(T-k)}) & k > (1-q)T - q \end{cases}.$$

The regret bounds for $q \in [\frac{1}{2}, 1]$ are derived from the Theorem 14 using the following lemma.

Lemma 15 *For every bit sequence $\Gamma = (\gamma_1, \ldots, \gamma_T)$ define $\bar{\Gamma} = (1-\gamma_1, \ldots, 1-\gamma_T)$. Then, $Regret^q_{TS(q)}(\Gamma) = Regret^{1-q}_{TS(1-q)}$*

The following theorem derives the worst-case sequences regret bound for general q.

Theorem 16 *For any observation sequence of length T, the regret of $TS(q)$ is $O\left(\sqrt{q(1-q)T}\right)$.*

6.3. Best-case regret bound

We do not characterize the exact best-case regret sequences[2], but only show that there are sequence with regret at most 1.

Theorem 17 *For every $q \in (0,1)$ and $m, n \in \mathbb{N}$, if $qm \leq (1-q)n$, then $Regret^q_{TS(q)}(1^n 0^m) \leq 1$ and otherwise $Regret^q_{TS(q)}(0^m 1^n) \leq 1$.*

7. Conclusion and further research

This paper studies Thompson sampling in an adversarial bit prediction setting. We give a full characterization for this particular environment, which enables us to understand the best and worst case behaviour. Our results show that TS has asymptomatically optimal results in the adversarial bit-prediction setting (i.e. $O(\sqrt{T})$ regret).

We also consider an extension for the bit prediction environment by adding weights to false positive and false negative mistakes. Using the same proof techniques, we managed to characterize this environment as well.

There are several natural directions for further research.

2. Finding the best-case sequence characterization for a general trade-off parameter q is harder than the previous cases. With the tools we presented, it is difficult even to compare the regrets of the bit sequences 10^k and 0^k1 for $k \in \mathbb{N}$.

- In order to get a full understanding of TS in an adversarial environment, we would want to extend the result to an adversarial MAB environment. This is important for a full-information environment as well as a partial one. We expect TS to be asymptotically as good as the known algorithms today.
- It would be interesting to find lower and upper bounds for the generalized bit-prediction environment. It is also interesting to compare our $O(\sqrt{q(1-q)T})$ regret bound to the performance of other known algorithms in this environment.

Acknowledgments

This work was supported in part by the Yandex Initiative in Machine Learning and by a grant from the Israel Science Foundation (ISF).

References

Shipra Agrawal and Navin Goyal. Further optimal regret bounds for thompson sampling. In *Artificial Intelligence and Statistics*, pages 99–107, 2013.

Shipra Agrawal and Navin Goyal. Near-optimal regret bounds for thompson sampling. *Journal of the ACM (JACM)*, 64(5):30, 2017.

Shipra Agrawal and Randy Jia. Optimistic posterior sampling for reinforcement learning: worst-case regret bounds. In *Advances in Neural Information Processing Systems*, pages 1184–1194, 2017.

Jean-Yves Audibert and Sébastien Bubeck. Minimax policies for adversarial and stochastic bandits. In *COLT*, pages 217–226, 2009.

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002a.

Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The nonstochastic multi-armed bandit problem. *SIAM journal on computing*, 32(1):48–77, 2002b.

Sébastien Bubeck and Nicolò Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.

Sébastien Bubeck and Mark Sellke. First-order regret analysis of thompson sampling. *arXiv preprint arXiv:1902.00681*, 2019.

Sébastien Bubeck and Aleksandrs Slivkins. The best of both worlds: Stochastic and adversarial bandits. In *The 25th Annual Conference on Learning Theory (COLT)*, pages 42.1–42.23, 2012.

N. Cesa-Bianchi and G. Lugosi. On prediction of individual sequences. *The Annals of Statistics*, 27 (6):1865—-1895, 1999.

Nicolò Cesa-Bianchi and Gabor Lugosi. *Prediction, learning, and games*. Cambridge University Press, 2006.

Olivier Chapelle and Lihong Li. An empirical evaluation of thompson sampling. In *Advances in neural information processing systems*, pages 2249–2257, 2011.

T.M. Cover. Behavior of sequential predictors of binary sequences. In *Transactions of the Fourth Prague Conference on Information Theory*, 1966.

Shagnik Das. A brief note on estimates of binomial coefficients. http://page.mi.fu-berlin.de/shagnik/notes/binomials.pdf.

DLMF. *NIST Digital Library of Mathematical Functions*. http://dlmf.nist.gov/, Release 1.0.15 of 2017-06-01. URL http://dlmf.nist.gov/. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller and B. V. Saunders, eds.

Meir Feder, Neri Merhav, and Michael Gutman. Universal prediction of individual sequences. *IEEE Trans. Information Theory*, 38(4):1258–1270, 1992.

Yoav Freund et al. Predicting a binary sequence almost as well as the optimal biased coin. Citeseer, 1996.

Aditya Gopalan. Thompson sampling for online learning with linear experts. *CoRR*, abs/1311.0468, 2013.

Nick Gravin, Yuval Peres, and Balasubramanian Sivan. Towards optimal algorithms for prediction with expert advice. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 528–547, 2016.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

Marcus Hutter. Optimality of universal bayesian sequence prediction for general loss and alphabet. *Journal of Machine Learning Research*, 4(Nov):971–1000, 2003.

Adam Tauman Kalai and Santosh Vempala. Efficient algorithms for online decision problems. *J. Comput. Syst. Sci.*, 71(3):291–307, 2005.

Emilie Kaufmann, Nathaniel Korda, and Rémi Munos. Thompson sampling: An asymptotically optimal finite-time analysis. In *International Conference on Algorithmic Learning Theory*, pages 199–213. Springer, 2012.

Jaya Kawale, Hung H Bui, Branislav Kveton, Long Tran-Thanh, and Sanjay Chawla. Efficient thompson sampling for online matrix-factorization recommendation. In *Advances in neural information processing systems*, pages 1297–1305, 2015.

Tor Lattimore and Csaba Szepesvári. Bandit algorithms. http://downloads.tor-lattimore.com/banditbook/book.pdf, 2019.

Neri Merhav and Meir Feder. Universal prediction. *IEEE Transactions on Information Theory*, 44 (6):2124–2147, 1998.

Michael Mitzenmacher and Eli Upfal. *Probability and computing: Randomized algorithms and probabilistic analysis*. Cambridge university press, 2005.

Jaouad Mourtada and Stephane Gaiffas. On the optimality of the hedge algorithm in the stochastic regime. *CoRR*, arXiv:1809.01382, 2018.

Ian Osband and Benjamin Van Roy. Why is posterior sampling better than optimism for reinforcement learning? In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 2701–2710. JMLR. org, 2017.

Ian Osband, Daniel Russo, and Benjamin Van Roy. (more) efficient reinforcement learning via posterior sampling. In *Advances in Neural Information Processing Systems*, pages 3003–3011, 2013.

Alexander Rakhlin and Karthik Sridharan. Statistical learning and sequential prediction. http://www.mit.edu/ rakhlin/courses/stat928/stat928_notes.pdf, 2014.

Daniel Russo and Benjamin Van Roy. An information-theoretic analysis of thompson sampling. *Journal of Machine Learning Research*, 17:68:1–68:30, 2016.

Daniel J Russo, Benjamin Van Roy, Abbas Kazerouni, Ian Osband, Zheng Wen, et al. A tutorial on thompson sampling. *Foundations and Trends® in Machine Learning*, 11(1):1–96, 2018.

Steven L Scott. A modern bayesian look at the multi-armed bandit. *Applied Stochastic Models in Business and Industry*, 26(6):639–658, 2010.

Yevgeny Seldin and Gábor Lugosi. A lower bound for multi-armed bandits with expert advice.

Yevgeny Seldin and Aleksandrs Slivkins. One practical algorithm for both stochastic and adversarial bandits. In *Proceedings of the 31th International Conference on Machine Learning*, pages 1287–1295, 2014.

Aleksandrs Slivkins. Introduction to multi-armed bandits. *arXiv preprint arXiv:1904.07272*, 2019.

William R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3–4):285–294, 1933.

Qun Xie and Andrew R Barron. Asymptotic minimax regret for data compression, gambling, and prediction. *IEEE Transactions on Information Theory*, 46(2):431–445, 2000.

Appendix A. Beta and Binomial concentration bounds

The following identities are well known (see, for example, Agrawal and Goyal (2017), Fact 3 and DLMF, Eq. 8.17.4).

The first relates the CDFs of the Beta and the Binomial distributions. The second is a property of the Beta distribution.

Fact 18 *For $a, b \in \mathbb{N}^+$ and $p \in [0,1]$ we have $F_{\beta(a,b)}(p) = 1 - F_{Bin(a+b-1,p)}(a-1)$.*

Fact 19 *For $a \in \mathbb{N}^+$ and $p \in [0,1]$ we have $F_{\beta(a,1)}(p) = p^a$.*

Next, we present concentration bounds and inequalities that we need for our proofs.

Fact 20 *(Gaussian Half CDF)*
Let $\sigma \in \mathbb{R}^+$. Then $\frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty e^{-\frac{x^2}{2\sigma^2}} dx = \frac{1}{2}$.

Fact 21 *(Multiplicative Chernoff bound) Mitzenmacher and Upfal (2005)*
Let $X_1, ..., X_n$ be random variables with values of $\{0,1\}$ such that $\mathrm{E}[X_t | X_1, ..., X_{t-1}] = \mu$. Let $S_n = \sum_{i=1}^{n} X_i$.

1. *For $1 \geq a \geq 0$, $\Pr(S_n \geq (1+a)n\mu) \leq e^{-\frac{a^2 n\mu}{3}}$.*
2. *For $a \geq 1$, $\Pr(S_n \geq (1+a)n\mu) \leq e^{-\frac{an\mu}{3}}$.*

Fact 22 *(Chernoff-Hoeffding) Hoeffding (1963)*
Let $X_1, ..., X_n$ be random variables with common range $[0,1]$ such that $\mathrm{E}[X_t | X_1, ..., X_{t-1}] = \mu$. Let $S_n = \sum_{i=1}^{n} X_i$.

1. *For all $a \geq 0$, $\Pr(|S_n - n\mu| \geq a) \leq 2e^{-\frac{2a^2}{n}}$.*
2. *For $\mu \geq \frac{1}{2}$ and $a \geq 0$, $\Pr(S_n > n\mu + a) \leq e^{-\frac{a^2}{2n\mu(1-\mu)}}$.*

Appendix B. Proof of bounds on sums of Beta CDFs (Theorems 3 and 25)

We present two bounds for sums of Beta CDFs. In the first subsection we prove a simple version of our bound, which appears Theorem 3. In the second subsection we expend the result to a general $q \in (0,1)$.

B.1. Proof of Theorem 3

The proof is divided into two parts. First we prove a bound on a series of exponents and then use Hoeffding bound to show that the exponent series is an upper bound for the sum of beta-distribution CDFs appears in Theorem 3.

Lemma 23 *For every $n \geq 1$, $\sum_{i=n+1}^{\infty} e^{-\frac{(i-(n+1))^2}{2(i+n+1)}} = \Theta(\sqrt{n})$.*

Proof Let $j = i - (n+1)$, then

$$\sum_{i=n+1}^{\infty} e^{-\frac{(i-(n+1))^2}{2(i+n+1)}} = \sum_{j=0}^{\infty} e^{-\frac{j^2}{2(j+2(n+1))}}. \tag{2}$$

We bound from below and above the exponents. For the upper bound we use the fact that $j \geq 0$ and for lower bounding the exponent we consider two cases: (a) $j > 2(n+1)$ and (b) $2(n+1) \geq j \geq 0$. We have,

$$\frac{j^2}{4(n+1)} \geq \frac{j^2}{2\left(j+2\left(n+1\right)\right)} \geq \begin{cases} \frac{j^2}{8(n+1)} & 2(n+1) \geq j \geq 0 \\ \frac{j}{4} & j > 2(n+1) \end{cases}.$$

We bound the sum (2) from below using Fact 20, where $\sigma^2 = 2(n+1)$, as follows

$$\sum_{j=0}^{\infty} e^{-\frac{j^2}{2(j+2(n+1))}} \geq \sum_{j=0}^{\infty} e^{-\frac{j^2}{4(n+1)}} \geq \sqrt{4\pi(n+1)}\frac{1}{\sqrt{4\pi(n+1)}}\int_0^{\infty} e^{-\frac{x^2}{4(n+1)}}dx = \sqrt{\pi(n+1)}.$$

For upper bounding Eq. (2) we have,

$$\sum_{j=0}^{\infty} e^{-\frac{j^2}{2(j+2(n+1))}} \leq \sum_{j=0}^{2(n+1)} e^{-\frac{j^2}{8(n+1)}} + \sum_{j=2(n+1)}^{\infty} e^{-\frac{j}{4}}. \tag{3}$$

The first sum of the right side of Eq. (3) is bounded, by using Fact 20 with $\sigma^2 = 4(n+1)$, as follows

$$\sum_{j=0}^{2(n+1)} e^{-\frac{j^2}{8(n+1)}} \leq 1 + \int_0^{2(n+1)} e^{-\frac{x^2}{8(n+1)}}dx \leq 1 + \sqrt{2\pi\left(n+1\right)}.$$

The second sum of the right hand side of Eq. (3) is an exponential sum and bounded as follows,

$$\sum_{j=2(n+1)}^{\infty} e^{-\frac{j}{4}} = \frac{1}{1-e^{-\frac{1}{4}}} - \frac{1-\left(e^{-\frac{1}{4}}\right)^{2n+3}}{1-e^{-\frac{1}{4}}} \leq \frac{1}{1-e^{-\frac{1}{4}}}.$$

By combining the previous inequalities and Eq. (3) we get $\sum_{i=n+1}^{\infty} e^{-\frac{(i-(n+1))^2}{2(i+n+1)}} = \Theta(\sqrt{n})$. ■

Theorem 3 *For every $n \geq 1$ we have $\sum_{i=n+1}^{\infty} F_{\beta(i+1,n+1)}\left(\frac{1}{2}\right) = O(\sqrt{n})$.*

Proof Using Fact 18

$$\begin{aligned}
\sum_{i=n+1}^{\infty} F_{\beta(i+1,n+1)}\left(\frac{1}{2}\right) &= \sum_{i=n+1}^{\infty}\left(1 - F_{Bin\left(i+n+1,\frac{1}{2}\right)}(i)\right) \\
&= \sum_{i=n+1}^{\infty}\left(1 - \Pr_{x_j \sim Ber\left(\frac{1}{2}\right)}\left(\sum_{j=1}^{i+n+1} x_j \leq i\right)\right) \\
&= \sum_{i=n+1}^{\infty} \Pr_{x_j \sim Ber\left(\frac{1}{2}\right)}\left(\sum_{j=1}^{i+n+1} x_j - \frac{i+n+1}{2} \geq \frac{i-(n+1)}{2}\right).
\end{aligned}$$

Note that $\frac{i-(n+1)}{2} \geq 0$ when $i \geq n+1$, therefore we can use the Chernoff-Hoffding bound (Fact 22.1) to achieve

$$\sum_{i=n+1}^{\infty} F_{\beta(i+1,n+1)}\left(\frac{1}{2}\right) \leq 2 \sum_{i=n+1}^{\infty} e^{-\frac{(i-(n+1))^2}{2(i+n+1)}} = \Theta(\sqrt{n}).$$

where the last equality follows from Lemma 23. ■

B.2. Proof of Theorem 25

The following subsection generalizes the proof of Theorem 3, as presented in Appendix B.1. We divide the generalized theorem version proof into two parts similarly to Appendix B.1.

Lemma 24 *For every $n \in \mathbb{N}^+$, $a > 0$ and $p \in (0,1)$ we have*

1. $$\sum_{i=\left\lceil \frac{p}{1-p}(n+1)\right\rceil+1}^{\infty} e^{-\frac{((1-p)i-p(n+1))^2}{a(i+n+1)}} \leq \frac{\sqrt{\pi a(n+1)}}{\sqrt{2}(1-p)^{3/2}} + \frac{2a}{(1-p)^2} e^{-\frac{1-p}{2a}(n+1)},$$

2. $$\sum_{i=\left\lfloor \frac{2p(n+1)}{1-2p}\right\rfloor+1}^{\infty} e^{-\frac{(1-p)i-p(n+1)}{a}} \leq 1 + \frac{a}{1-p} e^{-\frac{p(n+1)}{a(1-2p)}}.$$

Proof

1. We bound the sum as follows

$$\sum_{i=\left\lceil \frac{p}{1-p}(n+1)\right\rceil+1}^{\infty} e^{-\frac{((1-p)i-p(n+1))^2}{a(i+n+1)}} \leq \int_{\frac{p}{1-p}(n+1)}^{\infty} e^{-\frac{((1-p)x-p(n+1))^2}{a(x+n+1)}} dx.$$

Using a substitution of $y = (1-p)x - p(n+1)$,

$$\int_{\frac{p}{1-p}(n+1)}^{\infty} e^{-\frac{((1-p)x-p(n+1))^2}{a(x+n+1)}} dx \leq \frac{1}{1-p} \int_{0}^{\infty} e^{-\frac{y^2}{a(\frac{y+p(n+1)}{1-p}+n+1)}} dy = \frac{1}{1-p} \int_{0}^{\infty} e^{-\frac{1-p}{a(y+n+1)}y^2} dy. \quad (4)$$

We bound the exponent from below by considering two cases $y > n+1$ and $n+1 \geq y \geq 0$. We have,

$$\frac{1-p}{a(y+n+1)}y^2 \geq \begin{cases} \frac{1-p}{2a(n+1)}y^2 & n+1 \geq y \geq 0 \\ \frac{1-p}{2a}y & y > n+1 \end{cases}.$$

Hence, we have

$$\int_{0}^{\infty} e^{-\frac{1-p}{a(y+n+1)}y^2} dy \leq \int_{0}^{n+1} e^{-\frac{1-p}{2a(n+1)}y^2} dy + \int_{n+1}^{\infty} e^{-\frac{1-p}{2a}y} dy. \quad (5)$$

We bound the first integral of Eq. (5) using Fact 20, where $\sigma^2 = \frac{a(n+1)}{1-p}$, as follows

$$\int_0^{n+1} e^{-\frac{1-p}{2a(n+1)}y^2} dy \le \sqrt{\frac{2\pi a(n+1)}{1-p}} \sqrt{\frac{1-p}{2\pi a(n+1)}} \int_0^{\infty} e^{-\frac{1-p}{2a(n+1)}y^2} dy = \sqrt{\frac{\pi a(n+1)}{2(1-p)}}. \quad (6)$$

The second integral in Eq. (5) equals

$$\int_{n+1}^{\infty} e^{-\frac{1-p}{2a}y} dy = \frac{2a}{1-p} e^{-\frac{1-p}{2a}(n+1)}. \quad (7)$$

Combining Eq. (4 - 7) we have

$$\sum_{i=\left\lceil \frac{p}{1-p}(n+1) \right\rceil + 1}^{\infty} e^{-\frac{((1-p)i - p(n+1))^2}{a(i+n+1)}} \le \frac{\sqrt{\pi a(n+1)}}{\sqrt{2}(1-p)^{3/2}} + \frac{2a}{(1-p)^2} e^{-\frac{1-p}{2a}(n+1)}.$$

2. We bound the sum as follows

$$\sum_{i=\left\lfloor \frac{2p(n+1)}{1-2p} \right\rfloor + 1}^{\infty} e^{-\frac{(1-p)i - p(n+1)}{a}} \le 1 + \int_{\frac{2p(n+1)}{1-2p}}^{\infty} e^{-\frac{(1-p)x - p(n+1)}{a}} dx.$$

Using a substitution of $y = (1-p)x - p(n+1)$,

$$1 + \int_{\frac{2p(n+1)}{1-2p}}^{\infty} e^{-\frac{(1-p)x - p(n+1)}{a}} dx \le 1 + \frac{1}{1-p} \int_{\frac{p(n+1)}{1-2p}}^{\infty} e^{-\frac{y}{a}} dy = 1 + \frac{a}{1-p} e^{-\frac{p(n+1)}{a(1-2p)}}.$$

■

Theorem 25 *For every $n \ge 1$ and $p \in (0,1)$ we have*

$$\sum_{i=\left\lfloor \frac{p}{1-p}n \right\rfloor + 1}^{\infty} F_{\beta(i+1,n+1)}(p) = \begin{cases} 2\sqrt{3\pi p(n+1)} + O(1) & p \le \frac{1}{2} \\ 1 + \frac{p}{1-p} + \frac{\sqrt{\pi p(n+1)}}{1-p} + \frac{4p}{1-p} e^{-\frac{1}{4p}(n+1)} & p \ge \frac{1}{2} \end{cases}.$$

Proof Using Fact 18

$$\begin{aligned} \sum_{i=\left\lfloor \frac{p}{1-p}n \right\rfloor + 1}^{\infty} F_{\beta(i+1,n+1)}(p) &= \sum_{i=\left\lfloor \frac{p}{1-p}n \right\rfloor + 1}^{\infty} \left(1 - F_{Bin(i+n+1,p)}(i)\right) \quad (8) \\ &= \sum_{i=\left\lfloor \frac{p}{1-p}n \right\rfloor + 1}^{\infty} \left(1 - \Pr_{X_j \sim Ber(p)} \left(\sum_{j=1}^{i+n+1} X_j \le i \right)\right) \\ &= \sum_{i=\left\lfloor \frac{p}{1-p}n \right\rfloor + 1}^{\infty} \Pr_{X_j \sim Ber(p)} \left(\sum_{j=1}^{i+n+1} X_j > i \right). \end{aligned}$$

Let $N_i = i + n + 1$ and $r_i = (1-p)i - p(n+1)$. We have $i = pN_i + r_i$ and therefore, we rewrite Eq. (8) as

$$\sum_{i=\left\lfloor \frac{p}{1-p}n \right\rfloor + 1}^{\infty} F_{\beta(i+1,n+1)}(p) = \sum_{i=\left\lfloor \frac{p}{1-p}n \right\rfloor + 1}^{\infty} \Pr_{X_j \sim Ber(p)} \left(\sum_{j=1}^{N_i} X_j > pN_i + r_i \right). \tag{9}$$

1. First, we focus on the case of $p \leq \frac{1}{2}$.

Consider $\frac{r_i}{pN_i}$ and notice that $1 > \frac{r_i}{pN_i} \geq 0$ when $1 > \frac{(1-p)i - p(n+1)}{p(i+n+1)} \geq 0$, which is equivalent to $\frac{2p}{1-2p}(n+1) > i \geq \frac{p}{1-p}(n+1)$. Also, we note that $\mathrm{E}_{X_j \sim Ber(p)}[\sum_{j=1}^{N_i} X_j] = pN_i$. Using Chernoff bound (Fact 21.1) and Lemma 24.1, with $a = 3p$, we have

$$\begin{aligned} \sum_{i=\left\lceil \frac{p}{1-p}(n+1) \right\rceil + 1}^{\left\lfloor \frac{2p}{1-2p}(n+1) \right\rfloor} \Pr_{X_j \sim Ber(p)} \left(\sum_{j=1}^{N_i} X_j > pN_i + r_i \right) &\leq \sum_{i=\left\lceil \frac{p}{1-p}(n+1) \right\rceil + 1}^{\left\lfloor \frac{2p}{1-2p}(n+1) \right\rfloor} e^{-\frac{r_i^2}{3pN_i}} \\ &\leq \sum_{i=\left\lceil \frac{p}{1-p}(n+1) \right\rceil + 1}^{\infty} e^{-\frac{((1-p)i - p(n+1))^2}{3p(i+n+1)}} \leq \frac{\sqrt{3\pi p(n+1)}}{\sqrt{2}(1-p)^{3/2}} + \frac{6p}{(1-p)^2} e^{-\frac{1-p}{6p}(n+1)}. \end{aligned} \tag{10}$$

When $i > \frac{2p}{1-2p}(n+1)$ we use the second form of Chernoff bound (Fact 21.2), followed by Lemma 24.2, with $a = 3$, to have

$$\begin{aligned} \sum_{i=\left\lfloor \frac{2p}{1-2p}(n+1) \right\rfloor + 1}^{\infty} \Pr_{X_j \sim Ber(p)} \left(\sum_{j=1}^{N_i} X_j > pN_i + r_i \right) &\leq \sum_{i=\left\lfloor \frac{2p}{1-2p}(n+1) \right\rfloor + 1}^{\infty} e^{-\frac{r_i}{3}} \\ &= \sum_{i=\left\lfloor \frac{2p}{1-2p}(n+1) \right\rfloor + 1}^{\infty} e^{-\frac{(1-p)i - p(n+1)}{3}} \leq 1 + \frac{3}{1-p} e^{-\frac{p(n+1)}{3(1-2p)}}. \end{aligned} \tag{11}$$

When $\frac{p}{1-p}(n+1) > i$ we can assume worst-case to get

$$\sum_{i=\left\lfloor \frac{p}{1-p}n \right\rfloor + 1}^{\left\lceil \frac{p}{1-p}(n+1) \right\rceil} \Pr_{X_j \sim Ber(p)} \left(\sum_{j=1}^{N_i} X_j > pN_i + r_i \right) \leq 1 + \frac{p}{1-p}. \tag{12}$$

By substituting Eq. (10-12) in Eq. (9) we have

$$\begin{aligned} \sum_{i=\left\lfloor \frac{p}{1-p}n \right\rfloor + 1}^{\infty} F_{\beta(i+1,n+1)}(p) \leq 2 + \frac{p}{1-p} + \frac{\sqrt{3\pi p(n+1)}}{\sqrt{2}(1-p)^{3/2}} \\ + \frac{6p}{(1-p)^2} e^{-\frac{1-p}{6p}(n+1)} + \frac{3}{1-p} e^{-\frac{p(n+1)}{3(1-2p)}}. \end{aligned}$$

Since $p \leq \frac{1}{2}$, we have $\frac{1}{2} \leq 1-p$, thus

$$\sum_{i=\left\lfloor \frac{p}{1-p}n \right\rfloor+1}^{\infty} F_{\beta(i+1,n+1)}(p) = 2\sqrt{3\pi p(n+1)} + O(1).$$

2. Now, consider $p \geq \frac{1}{2}$. Assume $i \geq \frac{p}{1-p}(n+1)$ and therefore $r_i = (1-p)i - p(n+1) \geq pn + p - pn - p = 0$. Using Hoeffding bound (Fact 22.2) we get that

$$\Pr_{X_j \sim Ber(p)} \left(\sum_{j=1}^{N_i} X_j > pN_i + r_i \right) \leq e^{-\frac{r_i^2}{2p(1-p)N_i}}.$$

Thus, by using Lemma 24.1, with $a = 2p(1-p)$, we have

$$\begin{aligned} \sum_{i=\left\lceil \frac{p}{1-p}(n+1) \right\rceil+1}^{\infty} \Pr_{X_j \sim Ber(p)} \left(\sum_{j=1}^{N_i} X_j > pN_i + r_i \right) &\leq \sum_{i=\left\lceil \frac{p}{1-p}(n+1) \right\rceil+1}^{\infty} e^{-\frac{r_i^2}{2p(1-p)N_i}} \\ &\leq \frac{\sqrt{\pi p(n+1)}}{1-p} + \frac{4p}{(1-p)} e^{-\frac{1}{4p}(n+1)}. \end{aligned} \tag{13}$$

For $i \leq \frac{p}{1-p}(n+1)$ we assume the worst-case bound to get

$$\sum_{i=\left\lfloor \frac{p}{1-p}n \right\rfloor+1}^{\left\lceil \frac{p}{1-p}(n+1) \right\rceil} \Pr_{X_j \sim Ber(p)} \left(\sum_{j=1}^{N_i} X_j > pN_i + r_i \right) \leq 1 + \frac{p}{1-p}. \tag{14}$$

By substituting Eq. (13, 14) in Eq. (9) and using Lemma 24.1, with $a = 2p(1-p)$, to have

$$\sum_{i=\left\lfloor \frac{p}{1-p}n \right\rfloor+1}^{\infty} F_{\beta(i+1,n+1)}(p) \leq 1 + \frac{p}{1-p} + \frac{\sqrt{\pi p(n+1)}}{1-p} + \frac{4p}{(1-p)} e^{-\frac{1}{4p}(n+1)}.$$

■

Appendix C. Proof of the Swapping Lemma (Lemma 4)

We start with the following preliminary lemma that states the probability of an error for $TS(q)$ given a history.

Lemma 26 *Fix a bit sequence $\Gamma = (\gamma_1, \ldots, \gamma_T) \in \{0,1\}^T$. For any $t \in [T]$ we have,*

$$\Pr[\hat{\gamma}_t \neq \gamma_t \mid \Gamma] = E[\mathbb{I}\{\hat{\gamma}_t \neq \gamma_t\} | \Gamma] = \begin{cases} 1 - F_{\beta(O_{t-1}+1, Z_{t-1}+1)}(q) & \gamma_t = 0 \\ F_{\beta(O_{t-1}+1, Z_{t-1}+1)}(q) & \gamma_t = 1 \end{cases}$$

Proof At time t, algorithm $TS(q)$ samples $x_t \sim \beta\left(O_{t-1}+1, Z_{t-1}+1\right)$, and predicts $\hat{\gamma}_t = 1$ if $x_t > q$ and $\hat{\gamma}_t = 0$ if $x_t \leq q$. Thus, for the case of $\gamma_t = 0$,

$$\Pr\left(\hat{\gamma}_t \neq \gamma_t = 0\right) = \Pr\left(x_t > q\right) = 1 - F_{\beta(O_{t-1}+1, Z_{t-1}+1)}(q),$$

and for the case of $\gamma_t = 1$,

$$\Pr\left(\hat{\gamma}_t \neq \gamma_t = 1\right) = \Pr\left(x_t \leq q\right) = F_{\beta(O_{t-1}+1, Z_{t-1}+1)}(q).$$

■

Now we can prove the Swapping Lemma, which compares the regret of two sequences that differ by a single swap operation.

Lemma 27 (Swapping Lemma) *Fix a bit sequence $\Gamma = (\gamma_1, \ldots, \gamma_T) \in \{0,1\}^T$. For every t, such that $\gamma_t = 0$ and $\gamma_{t+1} = 1$, we have*

$$Regret^q_{TS(q)}(\Gamma) < Regret^q_{TS(q)}(Swap(\Gamma, t)) \iff \frac{q}{1-q} > \frac{O_{t-1}+1}{Z_{t-1}+1}.$$

For every t, such that $\gamma_t = 1$ and $\gamma_{t+1} = 0$, we have

$$Regret^q_{TS(q)}(\Gamma) < Regret^q_{TS(q)}(Swap(\Gamma, t)) \iff \frac{q}{1-q} < \frac{O_{t-1}+1}{Z_{t-1}+1}.$$

In addition,

$$Regret^q_{TS(q)}(\Gamma) = Regret^q_{TS(q)}(Swap(\Gamma, t)) \iff \frac{q}{1-q} = \frac{O_{t-1}+1}{Z_{t-1}+1}.$$

Proof We consider the difference between the regret of $TS(q)$ on the bit sequence Γ and the bit sequence $Swap(\Gamma, t)$. The two bit sequences differ only at locations t and $t+1$. Since the benchmark of a sequence depends only on the total number of zeros and ones in the sequence, the benchmarks on Γ and $Swap(\Gamma, t)$ are identical, i.e., $statis^q(\Gamma) = static^q(Swap(\Gamma, t))$. Therefore, the difference between the regrets is equals to the loss difference at time t and $t+1$.

Consider time $t \in [T]$ such that $\gamma_t = 0$ and $\gamma_{t+1} = 1$.We have,

$$\begin{aligned}
&Regret^q_{TS(q)}\left(\Gamma\right) - Regret^q_{TS(q)}\left(Swap(\Gamma, t)\right) \\
&= \sum_{t=1}^{T} E\left[\ell^q(\hat{\gamma}_t, \gamma_t) \mid \Gamma\right] - \sum_{t=1}^{T} E\left[\ell^q(\hat{\gamma}_t, \gamma_t) \mid Swap(\Gamma, t)\right] \\
&= E\left[\ell^q(\hat{\gamma}_t, \gamma_t) \mid \Gamma\right] + E\left[\ell^q(\hat{\gamma}_{t+1}, \gamma_{t+1}) \mid \Gamma\right] \\
&\quad - \left(E\left[\ell^q(\hat{\gamma}_t, \gamma_t) \mid Swap(\Gamma, t)\right] + E\left[\ell^q(\hat{\gamma}_{t+1}, \gamma_{t+1}) \mid Swap(\Gamma, t)\right]\right) \\
&= E\left[\ell^q(\hat{\gamma}_t, 0) \mid \Gamma\right] + E\left[\ell^q(\hat{\gamma}_{t+1}, 1) \mid \Gamma\right] \\
&\quad - \left(E\left[\ell^q(\hat{\gamma}_t, 1) \mid Swap(\Gamma, t)\right] + E\left[\ell^q(\hat{\gamma}_{t+1}, 0) \mid Swap(\Gamma, t)\right]\right) \\
&= q\left(1 - F_{\beta(O_{t-1}+1, Z_{t-1}+1)}(q)\right) + (1-q)F_{\beta(O_{t-1}+1, Z_{t-1}+2)}(q) \\
&\quad - \left((1-q)F_{\beta(O_{t-1}+1, Z_{t-1}+1)}(q) + q\left(1 - F_{\beta(O_{t-1}+2, Z_{t-1}+1)}(q)\right)\right) \\
&= (1-q)F_{\beta(O_{t-1}+1, Z_{t-1}+2)}(q) + qF_{\beta(O_{t-1}+2, Z_{t-1}+1)}(q) - F_{\beta(O_{t-1}+1, Z_{t-1}+1)}(q),
\end{aligned}$$

where we used Lemma 26 for the equality before last.

By Fact 2, we have the following recurrence relations:

$$F_{\beta(a+1,b)}(x) = F_{\beta(a,b)}(x) - \frac{x^a(1-x)^b}{a\mathrm{B}(a,b)} \text{ and } F_{\beta(a,b+1)}(x) = F_{\beta(a,b)}(x) + \frac{x^a(1-x)^b}{b\mathrm{B}(a,b)}.$$

where $B(a,b)$ is the Beta function. Therefore,

$$\begin{aligned}
&Regret^q_{TS(q)}(\Gamma) - Regret^q_{TS(q)}(Swap(\Gamma,t)) \\
&= (1-q)F_{\beta(O_{t-1}+1,Z_{t-1}+2)}(q) + qF_{\beta(O_{t-1}+2,Z_{t-1}+1)}(q) - F_{\beta(O_{t-1}+1,Z_{t-1}+1)}(q) \\
&= (1-q)\left(F_{\beta(O_{t-1}+1,Z_{t-1}+1)}(q) + \frac{q^{O_{t-1}+1}(1-q)^{Z_{t-1}+1}}{(Z_{t-1}+1)\,B\,(O_{t-1}+1,Z_{t-1}+1)}\right) \\
&\quad + q\left(F_{\beta(O_{t-1}+1,Z_{t-1}+1)} - \frac{q^{O_{t-1}+1}(1-q)^{Z_{t-1}+1}}{(O_{t-1}+1)B(O_{t-1}+1,Z_{t-1}+1)}\right) \\
&\quad - F_{\beta(O_{t-1}+1,Z_{t-1}+1)} \\
&= \frac{q^{O_{t-1}+1}(1-q)^{Z_{t-1}+1}}{B\,(O_{t-1}+1,Z_{t-1}+1)}\left(\frac{1-q}{Z_{t-1}+1} - \frac{q}{O_{t-1}+1}\right),
\end{aligned} \tag{15}$$

We now analyse the $sign$ of the terms in Eq. (15). Since $\frac{q^{O_{t-1}+1}(1-q)^{Z_{t-1}+1}}{B(O_{t-1}+1,Z_{t-1}+1)} > 0$,

$$sign\Big(Regret^q_{TS(q)}\,(\Gamma) - Regret^q_{TS(q)}\,(Swap\,(\Gamma,t))\Big) = sign\left(\frac{(1-q)}{Z_{t-1}+1} - \frac{q}{O_{t-1}+1}\right).$$

Thus,

$$Regret^q_{TS(q)}\,(\Gamma) < Regret^q_{TS(q)}\,(Swap(\Gamma,t)) \iff \frac{q}{1-q} > \frac{O_{t-1}+1}{Z_{t-1}+1},$$

and equality holds iff $\frac{q}{1-q} = \frac{O_{t-1}+1}{Z_{t-1}+1}$.

The second case, where $\gamma_t = 1$ and $\gamma_{t+1} = 0$, is similar. ■

Appendix D. Worst-case regret proofs for $q = \frac{1}{2}$ (Section 5.1)

Consider bit sequences $\Gamma = (\gamma_1, \ldots, \gamma_T)$ with k zeros, where $k \leq \frac{T}{2}$ zeros. We first show that among these bit sequences the ones of largest regret are of the form $\{01,10\}^k 1^{T-2k}$. Then, we prove that the regret of each of these sequences is $\Theta(\sqrt{k})$.

Theorem 5 *For any $\Gamma_1, \Gamma_2 \in \{01,10\}^k 1^{T-2k}$ we have $Regret^{1/2}_{TS(\frac{1}{2})}(\Gamma_1) = Regret^{1/2}_{TS(\frac{1}{2})}(\Gamma_2)$. In addition, for any $\Gamma_3 \notin \{01,10\}^k 1^{T-2k}$ we have $Regret^{1/2}_{TS(\frac{1}{2})}(\Gamma_1) > Regret^{1/2}_{TS(\frac{1}{2})}(\Gamma_3)$.*

Given the above theorem, to bound the worst case regret of $TS(\frac{1}{2})$, we can focus on the sequence $W^k_T = \{01\}^k 1^{T-2k}$ and bound $Regret^{1/2}_{TS(\frac{1}{2})}(W^k_T)$.

Theorem 6 *For every $T \in \mathbb{N}^+$ and $k \leq \frac{T}{2}$ we have, $Regret^{1/2}_{TS(\frac{1}{2})}(W^k_T) = \Theta(\sqrt{k})$.*

Proof Let $W_T^k = (w_1, \ldots, w_T)$, where we have: (1) $w_t = 0$ for $t \in A_1 = \{2i-1 \mid i \in [k]\}$, (2) $w_t = 1$ for $t \in A_2 = \{2i \mid i \in [k]\}$, and (3) $w_t = 1$ for $t \in A_3 = \{i \mid i \geq 2k+1\}$. We bound the expected number of errors made by $TS(\frac{1}{2})$ on each of these three subsets. Then, from these bounds we derive a bound on the loss and the regret.

The expected number of false positive errors in A_1: Note that the only errors at times $t \in A_1$ are false positive since $w_t = 0$ for these t's. For $t \in A_1$ we have that $t = 2i-1$, and $O_{t-1} = Z_{t-1} = i-1$. Hence the algorithm $TS(\frac{1}{2})$ predicts $\hat{\gamma}_t = 0$ and $\hat{\gamma}_t = 1$ each with probability of $\frac{1}{2}$ and

$$\mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq w_t\} \mid W_T^k\right] = \frac{1}{2}.$$

When we sum over $t \in A_1$, we have

$$\sum_{t=1}^{k} \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq w_t\} \mid W_T^k\right] = \frac{k}{2}.$$

The expected number of false negative errors in A_2: Note that the only errors at times $t \in A_2$ are false negatives since $w_t = 1$. For $t \in A_2$ we have $t = 2i$, and $O_{t-1} = i-1$ and $Z_{t-1} = i$. By Lemma 26 and Fact 18 we have

$$\mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq w_t\} \mid W_T^k\right] = F_{\beta(i,i+1)}\left(\frac{1}{2}\right) = 1 - F_{Bin(2i,\frac{1}{2})}(i-1).$$

We can bound $F_{Bin(2i,\frac{1}{2})}(i-1)$ using Fact 32, in the following way

$$\begin{aligned} F_{Bin(2i,\frac{1}{2})}(i-1) &= \Pr_{X \sim Bin(2i,\frac{1}{2})}(X \leq i) - \Pr_{X \sim Bin(2i,\frac{1}{2})}(X = i) \\ &= \frac{1}{2} - (1+o(1))\frac{1}{\sqrt{\pi i}} \end{aligned}$$

Summing over $t \in A_2$ we have,

$$\sum_{i=1}^{k} \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_{2i} \neq w_{2i}\} \mid W_T^k\right] = \frac{k}{2} + \sum_{i=1}^{k}(1+o(1))\frac{1}{\sqrt{\pi i}} = \frac{k}{2} + \Theta(\sqrt{k})$$

The expected number of false negative in A_3: Note that the only errors at times $t \in A_3$ are false negative since $w_t = 1$ for these t's. For any $t \in A_3$ we have $Z_t = k$. Therefore,

$$\mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq w_t\} \mid W_T^k\right] = F_{\beta(t-k+1,k+1)}\left(\frac{1}{2}\right).$$

From Theorem 3 we have

$$\sum_{t=2k+1}^{T} \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq w_t\} \mid W_T^k\right] = \sum_{t=2k+1}^{T} F_{\beta(t-k+1,k+1)}\left(\frac{1}{2}\right) = O(\sqrt{k}).$$

Summing up the errors over A_1, A_2, and A_3 we get that the total number of errors is

$$\sum_{t=1}^{T} \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq w_t\} \mid W_T^k\right] = \frac{k}{2} + \left(\frac{k}{2} + \Theta(\sqrt{k})\right) + O(\sqrt{k}) = k + \Theta(\sqrt{k})$$

Recall that the regret is the total loss minus the best static bit prediction. Since we assume that $k \leq \frac{T}{2}$ it is equal to

$$Regret^{1/2}_{TS(\frac{1}{2})}(W_T^k) = \frac{1}{2}\sum_{t=1}^{T} \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq w_t\} \mid W_T^k\right] - \frac{1}{2}\min\{T-k, k\} = \Theta(\sqrt{k}).$$

■

Appendix E. Best-case regret proofs for $q = \frac{1}{2}$ (Section 5.2)

We show that for $k \leq \frac{T}{2}$, the lowest regret is for the bit sequence $B_T^k = 1^{T-k}0^k$. Then, we prove that its regret is $O(1)$ for any $k \leq \frac{T}{2}$.

Lemma 28 *For any $\Phi \in \{0,1\}^{T-2m}$, $Regret^{1/2}_{TS(\frac{1}{2})}(0^m1^m\Phi) = Regret^{1/2}_{TS(\frac{1}{2})}(1^m0^m\Phi)$.*

Proof Let $\Gamma^1 = (\gamma_1^1, \ldots, \gamma_T^1) = (0^m1^m, \Phi)$ and $\Gamma^2 = (\gamma_1^2, \ldots, \gamma_T^2) = (1^m0^m, \Phi)$. We show, using Lemma 26, that for each $t \in [T]$, we have $\mathrm{E}[\mathbb{I}\{\hat{\gamma}_t = \gamma_t^1\} \mid \Gamma^1] = \mathrm{E}[\mathbb{I}\{\hat{\gamma}_t = \gamma_t^2\} \mid \Gamma^2]$, which implies that Γ^1 and Γ^2 have the same expected loss. Since static bit prediction also has the same loss on Γ^1 and Γ^2 then they have the same regret.

For $t \leq m$, by Fact 1, we have

$$\mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t = \gamma_t^1\} \mid \Gamma^1\right] = 1 - F_{\beta(1,i+1)}\left(\frac{1}{2}\right) = F_{\beta(i+1,1)}\left(\frac{1}{2}\right) = \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t = \gamma_t^2\} \mid \Gamma^2\right].$$

For $m < t \leq 2m$ we have,

$$\mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t = \gamma_t^1\} \mid \Gamma^1\right] = F_{\beta(i+1,m+1)}\left(\frac{1}{2}\right) = 1 - F_{\beta(m+1,i+1)}\left(\frac{1}{2}\right) = \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t = \gamma_t^2\} \mid \Gamma^2\right].$$

For $t > 2m$ we have $O_t(\Gamma^1) = O_t(\Gamma^2)$ and $Z_t(\Gamma^1) = Z_t(\Gamma^2)$ and thus $\mathrm{E}[\mathbb{I}\{\hat{\gamma}_t = \gamma_t^1\} \mid \Gamma^1] = \mathrm{E}[\mathbb{I}\{\hat{\gamma}_t = \gamma_t^2\} \mid \Gamma^2]$. ■

From that we can induce that B_T^k has the lowest regret on $TS(q)$.

Theorem 9 *The bit sequence with the lowest regret of length T with $k < \frac{T}{2}$ zeros is $B_T^k = 1^{T-k}0^k$. For $k = \frac{T}{2}$, both $1^{T/2}0^{T/2}$ and $0^{T/2}1^{T/2}$ have the lowest regret.*

Proof Let $\Gamma = (\gamma_1, \ldots, \gamma_T) \in \{0,1\}^T$ be a bit sequence of length T with $k \leq \frac{T}{2}$ zeros such that $\Gamma \neq 1^{T-k}0^k$. We show that there is a bit sequence $\tilde{\Gamma}$, that has the same regret as Γ, and for some $t \in [T]$ the sequence $Swap(\tilde{\Gamma}, t)$ has regret smaller than $\tilde{\Gamma}$.

Since $\Gamma \neq 1^{T-k}0^k$, then either $\Gamma = 0^k1^{T-k}$ or it has a prefix of the form 0^m1^n0 or 1^n0^m1, where $n, m > 0$.

First, we look at the case where $\Gamma = 0^k1^{T-k}$. By Lemma 28, the sequence $\tilde{\Gamma} = 1^k0^k1^{T-2k}$ has the same regret as Γ and by Lemma 4, the sequence $Swap(\tilde{\Gamma}, 2k)$ has regret smaller than the regret of $\tilde{\Gamma}$.

Second, assume Γ has a prefix of 0^m1^n0 (the case of 1^n0^m1 is similar). We have two subcases: (a) If $m \geq n$ then $O_{n+m-1} < Z_{n+m-1}$ and $\gamma_{n+m} = 1$, $\gamma_{n+m+1} = 0$. By Lemma 4, the sequence $Swap(\Gamma, n+m)$ has regret lower than Γ. (b) If $m < n$, by Lemma 28, the bit sequences $\Gamma = (0^m1^m1^{n-m}0, \gamma_{m+n+2}, \ldots, \gamma_T)$ and $\tilde{\Gamma} = (1^m0^m1^{n-m}0, \gamma_{m+n+2}, \ldots, \gamma_T)$ have the same regret. By Lemma 4, the sequence $Swap(\tilde{\Gamma}, 2m)$ has regret smaller than the regret of $\tilde{\Gamma}$.

For $k = \frac{T}{2}$, by Lemma 28, both $0^{T/2}1^{T/2}$ and $1^{T/2}0^{T/2}$ have the same regret. ■

We now bound the regret of $B_T^k = 1^{T-k}0^k$.

Theorem 10 *For every $T \in \mathbb{N}^+$ and $k \leq \frac{T}{2}$ we have, $Regret^{1/2}_{TS(\frac{1}{2})}(B_T^k) \leq 1$, where $B_T^k = 1^{T-k}0^k$.*

Proof For $t \leq T-k$ we have $b_t = 1$. Thus

$$\mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq b_t\} \mid B_T^k\right] = F_{\beta(O_{t-1}+1, Z_{t-1}+1)}\left(\frac{1}{2}\right) = F_{\beta(t,1)}\left(\frac{1}{2}\right).$$

Using Fact 19, we have

$$\mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq b_t\} \mid B_T^k\right] = \left(\frac{1}{2}\right)^t.$$

This implies that the expected number of false negative errors, in steps $t \leq T-k$, is

$$\sum_{t=1}^{T-k} \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq b_t\} \mid B_T^k\right] = \sum_{t=1}^{T-k}\left(\frac{1}{2}\right)^t \leq 1.$$

For $t \geq T-k+1$ we can have at most k errors so

$$\sum_{t=T-k+1}^{T} \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq b_t\} \mid B_T^k\right] \leq k.$$

Therefore, the regret of $TS(\frac{1}{2})$ on B_T^k is bounded by

$$\begin{aligned} Regret^{1/2}_{TS(\frac{1}{2})}(B_T^k) &= \frac{1}{2}\sum_{t=1}^{T} \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq b_t\} \mid B_T^k\right] - \frac{1}{2}\min\{T-k, k\} \\ &\leq \frac{1}{2}(k+1) - \frac{1}{2}\min\{T-k, k\} \leq 1. \end{aligned}$$

■

Appendix F. Worst-case regret proofs for a general q (Sections 6.1 and 6.2)

Recall H^q,

$$\forall \Phi \in \{0,1\}^* : H^q(\Phi) = \begin{cases} \{0\} & \frac{O(\Phi)+1}{Z(\Phi)+1} > \frac{q}{1-q} \\ \{1\} & \frac{O(\Phi)+1}{Z(\Phi)+1} < \frac{q}{1-q} \\ \{0,1\} & \frac{O(\Phi)+1}{Z(\Phi)+1} = \frac{q}{1-q} \end{cases}, \tag{16}$$

where $O(\Phi)$ is the total number of 1s in Φ and $Z(\Phi)$ is the total number of 0s in Φ. For every sequence $\Gamma = (\gamma_1, \ldots, \gamma_T) \in \{0,1\}^T$ we define $p(\Gamma)$ to be the largest index t s.t. $\forall i \in [t] : \gamma_i \in H^q(\Gamma_{1:i-1})$, where $\Gamma_{1:n} = (\gamma_1, \ldots, \gamma_n)$. We call a bit sequence $\Gamma = (\gamma_1, \ldots, \gamma_T)$ a *worst-case sequence* if $\gamma_{p(\Gamma)+1} = \ldots = \gamma_T$. We define the subsequence $(\gamma_1, \ldots, \gamma_{p(\Gamma)})$ as the *head* of Γ and denote it $head(\Gamma)$ and the subsequence $(\gamma_{p(\Gamma)+1}, \ldots, \gamma_T)$ as the *tail* of Γ and denote it $tail(\Gamma)$.

For start, we want to bound the number of 0s and 1s in the head of a worst-case sequence.

Lemma 29 *Fix a worst-case sequence $\Gamma = (\gamma_1, \ldots, \gamma_T)$ and let $t \le p(\Gamma)$. Then, if $\gamma_t = 0$ then $(1-q)t \le Z_t \le (1-q)t + (1-q)$ and $qt - (1-q) \le O_t \le qt$, if $\gamma_t = 1$ then $(1-q)t - q \le Z_t \le (1-q)t$ and $qt \le O_t \le qt + q$.*

Proof The proof is by induction on t. For $t = 1$ and $q < \frac{1}{2}$ we have that $\frac{q}{1-q} < 1$ and therefore H^q of an empty sequence equals $\{0\}$. Thus, as $t \le p(\Gamma)$, we must place $\gamma_1 = 0$. In case of such sequence $(1-q) \le 1 \le 2(1-q)$ and $2q - 1 \le 0 \le q$.

By the induction hypothesis for both $\gamma_{t-1} = 0$ and $\gamma_{t-1} = 1$ we have, $(1-q)(t-1) - q \le Z_{t-1} \le (1-q)(t-1) + (1-q)$ and $q(t-1) - (1-q) \le O_{t-1} \le q(t-1) + q$.

Case 1 $\gamma_t = 0$. Since $t \le p(\Gamma)$, we have that $0 \in H^q(\Gamma_{1:t-1})$ and therefore $\frac{O_{t-1}+1}{Z_{t-1}+1} \ge \frac{q}{1-q}$. Since $O_{t-1} = O_t$ and $Z_{t-1} + 1 = Z_t$ we get that

$$\frac{O_t + 1}{Z_t} \ge \frac{q}{1-q}. \tag{17}$$

Since $Z_t + O_t = t$ we can substitute $Z_t = t - O_t$ in Eq. (17) and get that $O_t \ge qt - (1-q)$. Similarly by substituting $O_t = t - Z_t$ in Eq. (17) we get that $Z_t \le (1-q)t + (1-q)$. The upper bound on O_t and the lower bound on Z_t follow directly from our assumption: $Z_t = Z_{t-1} + 1 \ge (1-q)(t-1) - q + 1 = (1-q)t$ and $O_t = O_{t-1} \le q(t-1) + q = qt$.

Case 2 $\gamma_t = 1$. Since $t \le p(\Gamma)$, we have that $1 \in H^q(\Gamma_{1:t-1})$ and therefore $\frac{O_{t-1}+1}{Z_{t-1}+1} \le \frac{q}{1-q}$. Since $O_{t-1} + 1 = O_t$ and $Z_{t-1} = Z_t$ we get that

$$\frac{O_t}{Z_t + 1} \le \frac{q}{1-q}. \tag{18}$$

Since $Z_t + O_t = t$ we can substitute $Z_t = t - O_t$ in Eq. (18) and get that $O_t \le qt + q$. Similarly by substituting $O_t = t - Z_t$ in Eq. (18) we get that $Z_t \ge (1-q)t - q$. The lower bound on O_t and the upper bound on Z_t follow directly from our assumption: $Z_t = Z_{t-1} \le (1-q)(t-1) + (1-q) = (1-q)t$ and $O_t = O_{t-1} + 1 \ge q(t-1) - (1-q) + 1 = qt$. ■

From Lemma 29 we characterize the tail of a worst-case sequence.

Theorem 11 *Let Γ be a worst-case sequence. If $Z_T \le (1-q)T - q$ then the $tail(\Gamma)$ is filled with ones. Otherwise, the $tail(\Gamma)$ is filled with zeros.*

Proof Let $j = p(\Gamma)$.

Consider first the case where $Z_T \leq (1-q)T - q$ and assume by contradiction that $tail(\Gamma)$ is not empty and it is filled with zeros. It follows from this assumption that $Z_T = Z_j + (T-j)$. By Lemma 29 we have that $Z_j \geq (1-q)j - q$, and by combining this inequality with the equality $Z_T = Z_j + (T-j)$ we get that $Z_T \geq (1-q)j - q + T - j = T - qj - q$. On the other hand we assumed that $Z_T \leq (1-q)T - q$. So by combining these upper and lower bounds on Z_T we get that $(1-q)T - q \geq T - qj - q$ and thus $j \geq T$. This is a contradiction to the assumption that $tail(\Gamma)$ is not empty.

Consider now the case where $Z_T \geq (1-q)T - q + 1$ and assume by contradiction that $tail(\Gamma)$ is not empty and it is filled with ones. It follows from this assumption that $O_T = O_j + (T-j)$. By Lemma 29 we have that $O_j \geq qj - (1-q)$, and by combining this inequality with the equality $O_T = O_j + (T-j)$ we get that $O_T \geq qj - (1-q) + T - j = T - (1-q)j - (1-q)$. On the other hand we assumed that $O_T = T - Z_T \leq qT - (1-q)$. So by combining these upper and lower bounds on O_T we get that $qT - (1-q) \geq T - (1-q)j - (1-q)$ and thus $j \geq T$. This is a contradiction to the assumption that $tail(\Gamma)$ is not empty. ■

Now we prove that all the worst-case sequences have the largest regret and bound it.

Theorem 12 *Let $\Gamma \in \{0,1\}^T$, s.t. Γ is not a worst-case sequence. Then, there exists $t \in [T]$ such that $Regret^q_{TS(q)}(\Gamma) < Regret^q_{TS(q)}(Swap(\Gamma, t))$.*

Proof Let $i = p(\Gamma) + 1$. Since Γ is not a worst-case sequence, there is an index $j > i$ such that $\gamma_j \neq \gamma_i$ (since $tail(\Gamma)$ contains both 0's and 1's). Assume j is the smallest index with this property.
Case 1 Assume $\gamma_i = 0$ and $\gamma_j = 1$. Since $\gamma_i \notin H^q(\Gamma_{1:i-1})$ we have $\frac{O_{i-1}(\Gamma)+1}{Z_{i-1}(\Gamma)+1} < \frac{q}{1-q}$. From the definition of j follows that $\gamma_i = \gamma_{i+1} = \ldots = \gamma_{j-1} = 0$ and thus $\frac{O_{j-2}(\Gamma)+1}{Z_{j-2}(\Gamma)+1} \leq \frac{O_{i-1}(\Gamma)+1}{Z_{i-1}(\Gamma)+1} < \frac{q}{1-q}$. By Lemma 4, the sequence $Swap(\Gamma, j-1)$ has a regret larger than Γ.
Case 2 Assume $\gamma_i = 1$ and $\gamma_j = 0$. Since $\gamma_i \notin H^q(\Gamma_{1:i-1})$ we have $\frac{O_{i-1}(\Gamma)+1}{Z_{i-1}(\Gamma)+1} > \frac{q}{1-q}$. From the definition of j follows that $\gamma_i = \gamma_{i+1} = \ldots = \gamma_{j-1} = 1$ and thus $\frac{O_{j-2}(\Gamma)+1}{Z_{j-2}(\Gamma)+1} \geq \frac{O_{i-1}(\Gamma)+1}{Z_{i-1}(\Gamma)+1} > \frac{q}{1-q}$. By Lemma 4, the sequence $Swap(\Gamma, j-1)$ has a regret larger than Γ. ■

Theorem 12 implies that any sequence of largest regret is a worst-case sequence. Next we prove that all worst-case sequences of length T with k zeros have the same regret.

Lemma 30 *All the worst-case sequences of length T with k zeros have the same regret.*

Proof Assume by contradiction that there are two worst-case sequences such that $Regret^q_{TS(q)}(\Gamma^1) = r_1$, $Regret^q_{TS(q)}(\Gamma^2) = r_2$ and $r_1 \neq r_2$. We assume further that Γ^1 and Γ^2 have the longest common prefix among all worst-case sequences of length T with k zeros and regret r_1 and r_2, respectively.

Since Γ^1 and Γ^2 both have k zeros then by Corollary 11 their tails are filled with the same bit. It follows that $head(\Gamma^1) \neq head(\Gamma^2)$. Assume without loss of generality that $head(\Gamma^2)$ is not shorter than $head(\Gamma^1)$. We claim that $head(\Gamma^1)$ is not a prefix of Γ^2. This follows since otherwise Γ^1 and Γ^2 cannot both have k zeros.

It follows that there exists an index $t \leq p(\Gamma^1)$ such that $\gamma^1_t \neq \gamma^2_t$. Let t be the smallest such index. Since $\Gamma^1_{1:t-1} = \Gamma^2_{1:t-1}$ we have that $\frac{O_{t-1}(\Gamma^1)+1}{Z_{t-1}(\Gamma^1)+1} = \frac{O_{t-1}(\Gamma^2)+1}{Z_{t-1}(\Gamma^2)+1} = \frac{q}{1-q}$. Assume that $\gamma^1_t = 0$

and $\gamma_t^2 = 1$. Therefore, there is an index $t' > t$ such that $\gamma_{t'}^1 = 1$ and $\gamma_{t'}^2 = 0$. Since the tails of both sequences are filled with the same bit then this implies that $t' \leq p(\Gamma^2)$ and therefore since $t+1 \leq t'$ we have that $t+1 \leq p(\Gamma^2)$.

Since $\gamma_t^2 = 1$ we have that $\frac{O_t(\Gamma^2)+1}{Z_t(\Gamma^2)+1} > \frac{O_{t-1}(\Gamma^2)+1}{Z_{t-1}(\Gamma^2)+1} = \frac{q}{1-q}$, and since $t+1 \leq p(\Gamma^2)$ we must have that $\gamma_{t+1}^2 = 0$. By Lemma 4, $Regret^q_{TS(q)}(\Gamma^2) = Regret^q_{TS(q)}\left(Swap(\Gamma^2, t)\right) = r_2$. It is easy to check that $Swap(\Gamma^2, t)$ is still a worst-case sequence and since it has a longer common prefix with Γ^1 we get a contradiction to the choice of Γ^1 and Γ^2.

The case where $\gamma_t^1 = 1$ and $\gamma_t^2 = 0$ is analogous. ■

Let $W_T^k = (w_1, \ldots, w_T) \in \{0,1\}^T$ be a worst-case sequence with k zeros such that for all $t \leq p(W_T^k)$ with $\frac{O_{t-1}+1}{Z_{t-1}+1} = \frac{q}{1-q}$ we have $\gamma_t = 0$. Since by Lemma 13 all the worst-case sequences with the same number of zeros have the same regret, we can focus on bounding the regret of W_T^k.

Theorem 14 *For every $T \in \mathbb{N}^+$, $q \in \left[0, \frac{1}{2}\right]$ and k zeros we have*

$$Regret^q_{TS(q)}(W_T^k) = \begin{cases} O(\sqrt{qk}) & k \leq (1-q)T - q \\ O(\sqrt{(1-q)(T-k)}) & k > (1-q)T - q \end{cases}.$$

Proof We first consider the case that $k \leq (1-q)T - q$. We partition W_T^k into the following sets (1) $A_1 = \{t \mid t \in [p(W_T^k)] \text{ and } w_t = 0\}$, (2) $A_2 = \{t \mid t \in [p(W_T^k)] \text{ and } w_t = 1\}$, and (3) $A_3 = \{t \mid t \geq p(W_T^k) + 1\}$. We bound the expected number of errors made by $TS(q)$ on each of these three subsets. Then, from these bounds we derive a bound on the loss and the regret.

The expected number of false positive errors in A_1: Note that the only errors at times $t \in A_1$ are false positive since $w_t = 0$ for these t's. Therefore, by Lemma 26 and Fact 18 we have

$$\begin{aligned} \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq w_t\} \mid W_T^k\right] &= 1 - F_{\beta(O_{t-1}+1, Z_{t-1}+1)}(q) = F_{\beta(Z_{t-1}+1, O_{t-1}+1)}(1-q) \\ &= 1 - F_{Bin(t,1-q)}(Z_{t-1}) = 1 - F_{Bin(t,1-q)}(Z_t - 1). \end{aligned} \tag{19}$$

By the definition of A_1, $t \leq p(W_T^k)$, and therefore by Lemma 29, $(1-q)t \leq Z_t$. Thus, $t \leq \frac{Z_t}{1-q} \leq \frac{Z_t+1-1+q}{1-q} = \frac{Z_t+1}{1-q} - 1 \leq \left\lfloor \frac{Z_t+1}{1-q} \right\rfloor$. Let $m = \left\lfloor \frac{Z_t+1}{1-q} \right\rfloor$ and $X \sim Bin\,(m, 1-q)$. We can bound the right side of Eq. (19) as follows.

$$\begin{aligned} F_{Bin(t,1-q)}(Z_t - 1) &\geq F_{Bin(m,1-q)}(Z_t - 1) \\ &= \Pr(X \leq Z_t + 1) - \Pr(X = Z_t + 1) - \Pr(X = Z_t). \end{aligned} \tag{20}$$

We now bound the different probabilities in Eq. (20). Since X is a Binomial random variable, its median is $\lfloor m(1-q) \rfloor = Z_t$ or $\lceil m(1-q) \rceil = Z_t + 1$ and thereby

$$\Pr(X \leq Z_t + 1) \geq \frac{1}{2}. \tag{21}$$

For any $Z_t \geq \frac{2(1-q)}{q} - 1$, we bound $\Pr(X = Z_t + 1)$ by Lemma 33 as follows

$$\Pr(X = Z_t + 1) = O\left(\frac{1}{\sqrt{qZ_t}}\right). \tag{22}$$

The probability $\Pr(X = Z_t)$ is bounded using the previous equality,

$$\frac{\Pr(X = Z_t)}{\Pr(X = Z_t + 1)} = \frac{q(Z_t + 1)}{(1-q)(m - Z_t)} \leq \frac{q(Z_t + 1)}{(1-q)(\frac{Z_t+1}{1-q} - 1 - Z_t)}$$
$$= \frac{q(Z_t + 1)}{(1-q)(\frac{Z_t+1-(1-q)(Z_t+1)}{1-q})} = \frac{q(Z_t + 1)}{q(Z_t + 1)} = 1. \tag{23}$$

Therefore by using Eq. (22) and (23) we have

$$\Pr(X = Z_t) = O\left(\frac{1}{\sqrt{qZ_t}}\right). \tag{24}$$

By substituting Eq. (20-22,24) into (19) we get that for $Z_t \geq \frac{2(1-q)}{q} - 1$

$$\mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq w_t\} \mid W_T^k\right] \leq \frac{1}{2} + O\left(\frac{1}{\sqrt{qZ_t}}\right). \tag{25}$$

For $Z_t < \frac{2(1-q)}{q} - 1$ we assume the worst-case to have

$$\mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq w_t\} \mid W_T^k\right] \leq 1. \tag{26}$$

Notice that since $k \leq (1-q)T - q$, by Corollary 11 there are no zeros in the tail.Thus, all the zeros of W_T^k are in A_1. Thus, we use Eq. (25-26) to sum over all $t \in A_1$.

$$\sum_{t \in A_1} \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq w_t\} \mid W_T^k\right]$$
$$= \sum_{\{t \in A_1 | Z_t < \frac{2(1-q)}{q} - 1\}} \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq w_t\} \mid W_T^k\right] + \sum_{\{t \in A_1 | Z_t \geq \frac{2(1-q)}{q} - 1\}} \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq w_t\} \mid W_T^k\right]$$
$$\leq \frac{2(1-q)}{q} - 1 + \sum_{t \in A_1} \left(\frac{1}{2} + O\left(\frac{1}{\sqrt{2\pi q Z_t}}\right)\right)$$
$$\leq O\left(\frac{1-q}{q}\right) + \frac{k}{2} + \sum_{i=1}^{k} O\left(\frac{1}{\sqrt{2\pi q i}}\right) = \frac{k}{2} + O\left(\sqrt{\frac{k}{q}} + \frac{1-q}{q}\right). \tag{27}$$

The expected number of false negative errors in A_2: Note that the only errors at times $t \in A_2$ are false negative since $w_t = 1$. Therefore, by Lemma 26 and Fact 18 we have

$$\mathrm{E}\left[I\{\hat{\gamma}_t \neq w_t\} \mid W_T^k\right] = F_{\beta(O_{t-1}+1, Z_{t-1}+1)}(q) = 1 - F_{Bin(t,q)}(O_{t-1}) = 1 - F_{Bin(t,q)}(O_t - 1). \tag{28}$$

By the definition of A_2, $t \leq p(W_T^k)$, and therefore by Lemma 29, $qt \leq O_t$. Thus, $t \leq \frac{O_t}{q} \leq \frac{O_t+1-q}{q} = \frac{O_t+1}{q} - 1 \leq \left\lfloor \frac{O_t+1}{q} \right\rfloor$. Let $m = \left\lfloor \frac{O_t+1}{q} \right\rfloor$ and $X \sim Bin(m, q)$. We can continue and bound the right side of Equation (28) as follows.

$$F_{Bin(t,q)}(O_t - 1) \geq F_{Bin(m,q)}(O_t - 1)$$
$$= \Pr(X \leq O_t + 1) - \Pr(X = O_t + 1) - \Pr(X = O_t). \tag{29}$$

Note that we have analogous bounds to the previous case of A_1, since by substituting Z_t and $1-q$ by O_t and q respectively in Eq. (19,20) we get Eq. (28,29). Thereby,

$$\mathrm{E}\left[I\{\hat{\gamma}_t \neq w_t\} \mid W_T^k\right] \leq \begin{cases} \frac{1}{2} + O\left(\frac{1}{\sqrt{(1-q)O_t}}\right) & O_t \geq \frac{2q}{1-q} - 1 \\ 1 & O_t < \frac{2q}{1-q} - 1 \end{cases}. \tag{30}$$

Since $head(W_T^k)$ contains all the zeros in W_T^k we have $Z_{p(W_T^k)} = k$. By using Lemma 29 we get that $(1-q)p(W_T^k) - q \leq Z_{p(W_T^k)}$ and thus $p(W_T^k) \leq \frac{k+q}{1-q}$. Therefore, $O_{p(W_T^k)} = p(W_T^k) - Z_{p(W_T^k)} \leq \frac{k+q}{1-q} - k \leq \frac{q}{1-q}k + 1$.

Let $n = \left\lceil \frac{q}{1-q}k \right\rceil + 1$. By Eq. (30), we sum over all $t \in A_2$ to have

$$\begin{aligned} \sum_{t \in A_2} \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq w_t\} \mid W_T^k\right] &\leq \frac{2q}{1-q} - 1 + \sum_{\{t \in A_2 \mid O_t \geq \frac{2q}{1-q}-1\}} O\left(\frac{1}{2} + \frac{1}{\sqrt{2\pi(1-q)O_t}}\right) \qquad (31) \\ &\leq \frac{2q}{1-q} - 1 + \frac{n}{2} + \sum_{i=1}^{n} O\left(\frac{1}{\sqrt{(1-q)i}}\right) = \frac{n}{2} + O\left(\sqrt{\frac{n}{1-q}} + \frac{q}{1-q}\right) \\ &\leq \frac{\frac{q}{1-q}k + 2}{2} + O\left(\sqrt{\frac{\frac{q}{1-q}k + 2}{1-q}} + \frac{q}{1-q}\right) = \frac{qk}{2(1-q)} + O\left(\frac{\sqrt{qk}}{1-q} + \frac{q}{1-q}\right), \end{aligned}$$

where the one before last inequality follows from substitution of $n = \left\lceil \frac{q}{1-q}k \right\rceil + 1 \leq \frac{q}{1-q}k + 2$.

The expected number of false negative in A_3: By Corollary 11 the only errors at times $t \in A_3$ are false negative since $w_t = 1$. For any $t \in A_3$ we have $Z_t = k$. Therefore,

$$\mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq w_t\} \mid W_T^k\right] = F_{\beta(t-k+1,k+1)}(q).$$

From Lemma 29, $(1-q)p(W_T^k) + (1-q) \geq Z_{p(W_T^k)} = k$ and thus $p(W_T^k) \geq \frac{k}{1-q} - 1$. From Theorem 25 we have

$$\begin{aligned} \sum_{t=\left\lfloor \frac{k}{1-q} \right\rfloor - 1}^{T} \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq w_t\} \mid W_T^k\right] &= \sum_{t=\left\lfloor \frac{k}{1-q} \right\rfloor - 1}^{T} F_{\beta(t-k+1,k+1)}(q) \leq \sum_{i=\left\lfloor \frac{qk}{1-q} \right\rfloor - 2}^{\infty} F_{\beta(i+1,k+1)}(q) \\ &\leq 3 + \sum_{i=\left\lfloor \frac{qk}{1-q} \right\rfloor + 1}^{\infty} F_{\beta(i+1,k+1)}(q) = O\left(\sqrt{qk}\right), \qquad (32) \end{aligned}$$

where the inequality follows from $t - k = \left\lfloor \frac{k}{1-q} \right\rfloor - 1 - k \geq \frac{k}{1-q} - k - 2 = \frac{qk}{1-q} - 2 = i$.

Since $k \leq (1-q)T - q$, the best static bit predictor is

$$static^q(W_T^k) = \min\{(1-q)(T-k), qk\} = qk.$$

By using Eq. (27), (31) and (32), the regret is the total loss minus the best static bit prediction

$$\begin{aligned}
Regret^q_{TS(q)}(W_T^k) &= \sum_{t=1}^{T} \mathrm{E}_{\hat{\gamma}_t \sim TS(q)} \left[\ell^q(\hat{\gamma}_t, w_t) \mid W_T^k \right] - static^q \left(W_T^k \right) \\
&= q \left(\frac{k}{2} + O\left(\sqrt{\frac{k}{q}} + \frac{1-q}{q} \right) \right) + (1-q) \left(\frac{qk}{2(1-q)} + O\left(\frac{\sqrt{qk}}{1-q} + \frac{q}{1-q} \right) \right) \\
&\quad + (1-q) O(\sqrt{qk}) - \min\left\{ (1-q)(T-k), qk \right\} \\
&= O\left(\sqrt{qk} \right).
\end{aligned}$$

We now look at the regret for $k \geq (1-q)T - q$. In this proof, we split the calculations into A_1, A_2 and A_3 as in the prior part.

The expected number of false positive errors in A_1: At each $t \in A_1$ the expected errors are bounded in the same way as in the previous case. The only change is the size of A_1. Notice that since $k > (1-q)T - q$, by Corollary 11 all ones of W_T^k are in A_1. By the definition of A_1, $t \leq p(W_T^k)$, and therefore by Lemma 29, $qt - (1-q) \leq O_{p(W_T^k)} = T - k$ and thus $t \leq \frac{T-k+1}{q}$. From Lemma 29 we also conclude that $Z_{p(W_T^k)} \leq (1-q)p(W_T^k) + (1-q) \leq (1-q)\frac{T-k+1}{q} + (1-q)$. In total, $|A_1| \leq (1-q)\frac{T-k+1}{q} + (1-q)$. Thereby, the expected number of errors in A_1 is bounded by $\frac{\frac{1-q}{q}(T-k)}{2} + O\left(\frac{\sqrt{(1-q)(T-k)}}{q} + \frac{1-q}{q} \right)$.

The expected number of false negative errors in A_2: At each $t \in A_2$ the expected errors are bounded in the same way as in the previous case. The only change is the size of A_2, which equals to $T-k$ since from Corollary 11 all the ones of W_T^k are in $head(W_T^k)$. Thus we have that the expected number of errors is bounded by $\frac{T-k}{2} + O\left(\sqrt{\frac{T-k}{1-q}} + \frac{q}{1-q} \right)$.

The expected number of false negative in A_3 : By Corollary 11 the only errors at times $t \in A_3$ are false negative since $w_t = 0$. For any $t \in A_3$ we have $O_t = T - k$. Therefore,

$$\mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq w_t\} \mid W_T^k \right] = 1 - F_{\beta(T-k+1, t-(T-k)+1)}(q) = F_{\beta(t-(T-k)+1, T-k+1)}(1-q).$$

From Lemma 29, $qp(W_T^k) + q \geq O_{p(W_T^k)} = T - k$ and thus $p(W_T^k) \geq \frac{T-k}{q} - 1$. From Theorem 25, since $1 - q \geq \frac{1}{2}$, we have

$$\begin{aligned}
\sum_{t=\left\lfloor \frac{T-k}{q} \right\rfloor - 1}^{T} \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq w_t\} \mid W_T^k \right] &= \sum_{t=\left\lfloor \frac{T-k}{q} \right\rfloor - 1}^{T} F_{\beta(t-(T-k)+1, T-k+1)}(1-q) \\
&\leq 3 + \sum_{i=\left\lfloor \frac{1-q}{q}(T-k) \right\rfloor + 1}^{\infty} F_{\beta(i+1, T-k+1)}(1-q) \\
&= O\left(\frac{\sqrt{(1-q)(T-k+1)}}{q} + \frac{1-q}{q} e^{-\frac{1}{4(1-q)}(T-k+1)} + \frac{1-q}{q} \right).
\end{aligned}$$

Since $k > (1-q)T - q$, the best static bit predictor is

$$static^q(W_T^k) = \min\{(1-q)(T-k), qk\} = (1-q)(T-k).$$

Hence, the regret in the case is

$$\begin{aligned}
Regret^q_{TS(q)}\left(W_T^k\right) &= \sum_{t=1}^{T} \mathrm{E}_{\hat{\gamma}_t \sim TS(q)}\left[\ell^q(\hat{\gamma}_t, w_t) \mid W_T^k\right] - static^q\left(W_T^k\right) \\
&= q\left(\frac{\frac{1-q}{q}(T-k)}{2} + O\left(\frac{\sqrt{(1-q)(T-k)}}{q} + \frac{1-q}{q}\right)\right) \\
&\quad + (1-q)\left(\frac{T-k}{2} + O\left(\sqrt{\frac{T-k}{1-q}} + \frac{q}{1-q}\right)\right) \\
&\quad + qO\left(\frac{\sqrt{(1-q)(T-k+1)}}{q} + \frac{1-q}{q}e^{-\frac{1}{4(1-q)}((T-k+1)+1)} + \frac{1-q}{q}\right) \\
&\quad - \min\{(1-q)(T-k), qk\} \\
&= O\left(\sqrt{(1-q)(T-k)}\right)
\end{aligned}$$

■

Lemma 31 *For every bit sequence* $\Gamma = (\gamma_1, \ldots, \gamma_T)$ *define* $\bar{\Gamma} = (1-\gamma_1, \ldots, 1-\gamma_T)$. *Then,* $Regret^q_{TS(q)}(\Gamma) = Regret^{1-q}_{TS(1-q)}$

Proof Fix $q \in \left[\frac{1}{2}, 1\right]$ and a bit sequence $\Gamma = (\gamma_1, \ldots, \gamma_T)$. We show that $Regret^q_{TS(q)}(\Gamma) = Regret^q_{TS(1-q)}\left(\bar{\Gamma}\right)$. At each step $t \in [T]$, $O_t(\Gamma) = Z_t\left(\bar{\Gamma}\right)$. Therefore by Fact 1 we have

$$\begin{aligned}
\mathrm{E}_{\hat{\gamma}_t \sim TS(q)}\left[\mathbb{I}\{\hat{\gamma}_{i_t} = 1\} \mid \Gamma\right] &= \Pr_{x_t \sim \beta(O_{t-1}(\Gamma)+1, Z_{t-1}(\Gamma)+1)}(x_t > q) \\
&= \Pr_{x_t \sim \beta(Z_{t-1}(\Gamma)+1, O_{t-1}(\Gamma)+1)}(x_t < 1-q) \\
&= \Pr_{x_t \sim \beta\left(O_{t-1}\left(\bar{\Gamma}\right)+1, Z_{t-1}\left(\bar{\Gamma}\right)+1\right)}(x_t < 1-q) \\
&= \mathrm{E}_{\hat{\gamma}_t \sim TS(1-q)}\left[\mathbb{I}\{\hat{\gamma}_{i_t} = 0\} \mid \bar{\Gamma}\right].
\end{aligned}$$

The benchmarks are the same as,

$$\begin{aligned}
static_q(\Gamma) &= \min\{(1-q)\, O_T(\Gamma), qZ_T(\Gamma)\} \\
&= \min\left\{qO_T\left(\bar{\Gamma}\right), (1-q)\, Z_T\left(\bar{\Gamma}\right)\right\} = static_{1-q}\left(\bar{\Gamma}\right).
\end{aligned}$$

We conclude that $Regret^q_{TS(q)}(\Gamma) = Regret^{1-q}_{TS(1-q)}\left(\bar{\Gamma}\right)$. ■

Theorem 16 *For any observation sequence of length* T, *the regret of* $TS(q)$ *is* $O\left(\sqrt{q(1-q)T}\right)$.

Proof Assume $q \in \left[0, \frac{1}{2}\right]$. From Theorem 12 the bit sequences that generate the largest regret, with k zeros, are worst-case sequences. Theorem 14 shows that the regret of these bit sequences is

$$\begin{cases} O\left(\sqrt{qk}\right) & k \leq (1-q)T - q \\ O\left(\sqrt{(1-q)(T-k)}\right) & \text{otherwise} \end{cases}.$$

Thus, the worst-case regret over all k's is

$$\max\left\{O\left(\sqrt{q(1-q)T}\right), O\left(\sqrt{(1-q)(T-(1-q)T)}\right)\right\} = O\left(\sqrt{q(1-q)T}\right).$$

For $q \in \left[\frac{1}{2}, 1\right]$, Lemma 15 with Theorem 14 gives us the same regret of $O\left(\sqrt{q(1-q)T}\right)$. ■

Appendix G. Best-case regret proofs for a general q (Section 6.3)

Theorem 17 *For every $q \in (0,1)$ and $m, n \in \mathbb{N}$, if $qm \leq (1-q)n$, then $Regret^q_{TS(q)}(1^n 0^m) \leq 1$ and otherwise $Regret^q_{TS(q)}(0^m 1^n) \leq 1$.*

Proof First we calculate the loss of $\Gamma_1 = 1^n 0^m$. For $t \leq n$ we have $\gamma_t = 1$. Thus, by using Lemma 26,

$$\mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq \gamma_t^{(1)}\} \mid \Gamma_1\right] = F_{\beta(O_{t-1}+1, Z_{t-1}+1)}(q) = F_{\beta(t,1)}(q).$$

Using Fact 19, we have

$$F_{\beta(t,1)}(q) = q^t.$$

This implies that the expected number of false negative errors, in steps $t \leq n$, is

$$\sum_{t=1}^{n} \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq \gamma_t^{(1)}\} \mid \Gamma_1\right] = \sum_{t=1}^{n} q^t \leq \frac{1}{1-q}.$$

For $t \geq n+1$ we can have at most m errors so

$$\sum_{t=n+1}^{T} \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq \gamma_t^{(1)}\} \mid \Gamma_1\right] \leq m.$$

Therefore, the expected loss of $TS(q)$ on Γ_1 is bounded by

$$\begin{aligned} \sum_{t=1}^{T} E\left[\ell^q(\hat{\gamma}_t, \gamma_t^{(1)}) \mid \Gamma_1\right] &= (1-q)\sum_{t=1}^{n} \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq \gamma_t^{(1)}\} \mid \Gamma_1\right] + q\sum_{t=n+1}^{n+m} \mathrm{E}\left[\mathbb{I}\{\hat{\gamma}_t \neq \gamma_t^{(1)}\} \mid \Gamma_1\right] \\ &\leq (1-q)\frac{1}{1-q} + qm = 1 + qm. \end{aligned} \tag{33}$$

Analogously, we bound the expected loss of $TS(q)$ on $\Gamma_2 = 0^m 1^n$ by

$$\sum_{t=1}^{T} E\left[\ell^q(\hat{\gamma}_t, \gamma_t^{(2)}) \mid \Gamma_2\right] \leq 1 + (1-q)n. \tag{34}$$

The benchmark of the two sequences is the same and equals

$$static^q(\Gamma_1) = static^q(\Gamma_2) = \min\{qm, (1-q)n\}.$$

Therefore, if $\min\{qm, (1-q)n\} = qm$ then by Eq. (33)

$$Regret^q_{TS(q)}(\Gamma_1) \leq 1 + qm - qm = 1.$$

Otherwise $\min\{qm, (1-q)n\} = (1-q)n$ and by Eq. (34)

$$Regret^q_{TS(q)}(\Gamma_2) \leq 1 + (1-q)n - (1-q)n = 1.$$

■

Appendix H. Binomial coefficient approximations

We use the following well known approximation of the Binomial coefficient using Stirling's approximation. (see for example, Das)

Fact 32 *For every $m \in \mathbb{N}^+$ and $n \leq m$ we have*

$$\binom{m}{n} = (1+o(1))\sqrt{\frac{m}{2\pi n(m-n)}}\left(\frac{m}{n}\right)^n\left(\frac{m}{m-n}\right)^{m-n}.$$

From the fact above we conclude the following lemma.

Lemma 33 *For every constant $p \in (0,1)$ and $n \geq \frac{2p}{1-p}$, we have*

$$\Pr_{X\sim Bin(\left\lfloor \frac{n}{p}\right\rfloor,p)}(X=n) = O\left(\frac{1}{\sqrt{(1-p)n}}\right).$$

Proof Let $m = \left\lfloor \frac{n}{p}\right\rfloor$. We bound $\binom{m}{n}$ using Fact 32 as follows

$$\binom{m}{n} = (1+o(1))\sqrt{\frac{m}{2\pi n(m-n)}}\left(\frac{m}{n}\right)^n\left(\frac{m}{m-n}\right)^{m-n}.$$

From the definition of floor $\exists \omega \in [0,1) : m = \frac{n}{p} - \omega$ and therefore

$$\begin{aligned}
\binom{m}{n} &= (1+o(1))\sqrt{\frac{\frac{n}{p}-\omega}{2\pi n(\frac{n}{p}-\omega-n)}}\left(\frac{\frac{n}{p}-\omega}{n}\right)^n\left(\frac{\frac{n}{p}-\omega}{\frac{n}{p}-\omega-n}\right)^{\frac{n}{p}-\omega-n} \\
&= O(1)\sqrt{\frac{\frac{n-p\omega}{p}}{n(\frac{(1-p)n-p\omega}{p})}}\left(\frac{\frac{n-p\omega}{p}}{n}\right)^n\left(\frac{\frac{n-p\omega}{p}}{\frac{(1-p)n-p\omega}{p}}\right)^{\frac{n}{p}-\omega-n} \\
&= O(1)\sqrt{\frac{n-p\omega}{n((1-p)n-p\omega)}}\left(\frac{n-p\omega}{pn}\right)^n\left(\frac{n-p\omega}{(1-p)n-p\omega}\right)^{\frac{n}{p}-\omega-n}.
\end{aligned}$$

Since $0 \leq p\omega < p$ we have

$$\begin{aligned}\binom{m}{n} &\leq O(1)\sqrt{\frac{n}{n((1-p)n-p)}}\left(\frac{n}{pn}\right)^n\left(\frac{n}{(1-p)n-p}\right)^{\frac{n}{p}-\omega-n} \\ &= O(1)\sqrt{\frac{1}{(1-p)n-p}}\left(\frac{1}{p}\right)^n\left(\frac{n}{(1-p)n-p}\right)^{\frac{n}{p}-\omega-n}. \end{aligned} \tag{35}$$

Since $n \geq \frac{2p}{1-p}$ we get that $\frac{(1-p)n}{2} \geq p$ and therefore $(1-p)n - p \geq \frac{(1-p)n}{2}$. Thus, by using Eq. (35),

$$\binom{m}{n} \leq O(1)\sqrt{\frac{2}{(1-p)n}}\left(\frac{1}{p}\right)^n\left(\frac{n}{(1-p)n-p}\right)^{\frac{n}{p}-\omega-n}. \tag{36}$$

We bound $\left(\frac{n}{(1-p)n-p}\right)^{\frac{n}{p}-\omega-n}$ as follow

$$\begin{aligned}\left(\frac{n}{(1-p)n-p}\right)^{\frac{n}{p}-\omega-n} &= (1-p)^{-(\frac{n}{p}-\omega-n)}\left(\frac{(1-p)n}{(1-p)n-p}\right)^{\frac{n}{p}-\omega-n} \\ &= (1-p)^{-(\frac{n}{p}-\omega-n)}\left(\frac{1}{1-\frac{p}{(1-p)n}}\right)^{\frac{n}{p}-\omega-n} \\ &\leq (1-p)^{-(\frac{n}{p}-\omega-n)}\frac{1}{\left(1-\frac{p}{(1-p)n}\right)^{\frac{(1-p)n}{p}}} \leq 4(1-p)^{-(m-n)}. \end{aligned} \tag{37}$$

where the last inequality holds as $\left(1-\frac{p}{(1-p)n}\right)^{\frac{(1-p)n}{p}}$ is a monotonic increasing function and since $n \geq \frac{2p}{1-p}$, the function has a minimum at $n = \frac{2p}{1-p}$.

From Eq. (36,37) we have

$$\Pr_{X\sim Bin(m,p)}(X=n) = \binom{m}{n}p^n(1-p)^{m-n} = O\left(\frac{1}{\sqrt{(1-p)n}}\right).$$

■