k-Cut: A Simple Approximately-Uniform Method for Sampling Ballots in Post-Election Audits*

Mayuri Sridhar and Ronald L. Rivest

Massachusetts Institute of Technology, Cambridge MA 02139, USA
mayuri@mit.edu
rivest@csail.mit.edu

Abstract. We present an approximate sampling framework and discuss how risk-limiting audits can compensate for these approximations, while maintaining their "risk-limiting" properties. Our framework is general and can compensate for counting mistakes made during audits.
Moreover, we present and analyze a simple approximate sampling method, "k-cut", for picking a ballot randomly from a stack, without counting. Our method involves doing k "cuts," each involving moving a random portion of ballots from the top to the bottom of the stack, and then picking the ballot on top. Unlike conventional methods of picking a ballot at random, k-cut does not require identification numbers on the ballots or counting many ballots per draw. We analyze how close the distribution of chosen ballots is to the uniform distribution, and design mitigation procedures. We show that $k = 6$ cuts is enough for a risk-limiting election audit, based on empirical data, which provides a significant increase in sampling efficiency. This method has been used in pilot RLAs in Indiana and is scheduled to be used in Michigan pilot audits in December 2018.

Keywords: sampling · elections · auditing · post-election audits · risk-limiting audit · Bayesian audit.

1 Introduction

The goal of post-election tabulation audits is to provide assurance that the reported results of the contest are correct; that is, they agree with the results that a full hand-count would reveal. To do this, the auditor draws ballots uniformly at random one at a time from the set of all cast paper ballots, until the sample of ballots provides enough assurance that the reported outcomes are correct.

The most popular post-election audit method is known as a "risk-limiting audit" (or RLA), invented by Stark (see his web page [13]). See also [3, 5–7, 11, 12] for explanations, details, and related papers. An RLA takes as input a "risk-limit" α (like 0.05), and ensures that if a reported contest outcome is incorrect, then this error will be detected and corrected with probability at least $1 - \alpha$.

* Supported by Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-0939370.

This paper provides a novel method for drawing a random sample of the cast paper ballots. The new method may often be more efficient than standard methods. However, it has a cost: ballots are drawn in a way that is only "approximately uniform". This paper provides ways of compensating for such non-uniformity.

There are two standard approaches for drawing a random sample of cast paper ballots:

1. [**ID-based sampling**] Print on each scanned cast paper ballot a unique identifying number (ballot ID numbers). Draw a random sample of ballot ID numbers, and retrieve the corresponding ballots.
2. [**Position-based sampling**] Give each ballot an implicit ballot ID equal to its position, then proceed as with method (1).

These methods work well, and are guaranteed to produce random samples. In practice, auditors use software, like [14], which takes in a ballot manifest as input and produces the random sample of ballot ID numbers. In this software, it is typically assumed that sampling is done without replacement.

However, finding even a single ballot using these sampling methods can be tedious and awkward in practice. For example, given a random sample of ID numbers, one may need to count or search through a stack of ballots to find the desired ballot with the right ID or at the right position. Moreover, typical auditing procedures assume that there are no mistakes when finding the ballots for the sample. Yet, this seems to be an unreasonable assumption - a study by Goggin et al. shows that when counting 120 ballots, human teams miscount the number of votes for a given candidate at an average rate of 1.4% [4]. In the literature about RLAs, there is no way to correct for these mistakes.

Our goal is to simplify the sampling process.

In particular, we define a general framework for compensating for "approximate sampling" in RLAs. Our framework of approximate sampling can be used to measure and compensate for human error rate while using the counting methods outlined above. Moreover, we also define a simpler approach for drawing a random sample of ballots, which does not rely on counting at all. Our technique is simple and easy to iterate on and may be of particular interest when the stack of ballots to be drawn from is large. We define mitigation procedures to account for the fact that the sampling technique is no longer uniformly random.

Overview of this paper. Section 2 introduces the relevant notation that we use throughout the paper.

Section 3 presents our proposed sampling method, called "k-cut."

Section 4 studies the distribution of single cut sizes, and provides experimental data. We then show how iterating a single cut provides improved uniformity for ballot selection.

Section 5 discusses the major questions that are brought up when using "approximate" sampling in a post-election audit.

Section 6 proves a very general result: that any general statistical auditing procedure for an arbitrary election can be adapted to work with approximate sampling, with simple mitigation procedures.

Section 7 discusses how to adapt the k-cut method for sampling when the ballots are organized into multiple stacks or boxes.

Section 8 provides some guidance for using k-cut in practice.

Section 9 gives some further discussion, lists some open problems, and makes some suggestions for further research.

Section 10 summarizes our contributions.

2 Notation and Election Terminology

Notation. We let $[n]$ denote the set $\{0, 1, \ldots, n-1\}$, and we let $[a, b]$ denote the set $\{a, a+1, \ldots, b-1\}$.

We let $\mathcal{U}[n]$ denote the uniform distribution over the set $[n]$. In $\mathcal{U}[n]$, the "$[n]$" may be omitted when it is understood to be $[n]$, where n is the number of ballots in the stack. We let $\mathcal{U}[a, b]$ denote the uniform distribution over the set $[a, b]$.

We let $VD(p, q)$ denote the variation distance between probability distributions p and q; this is the maximum, over all events E, of

$$Pr_p[E] - Pr_q[E].$$

Election Terminology. The term "ballot" here means to a single piece of paper on which the voter has recorded a choice for each contest for which the voter is eligible to vote. One may refer to a ballot as a "card." Multi-card ballots are not discussed in this paper.

Audit types. There are two kinds of post-election audits: *ballot-polling* audits, and *ballot-comparison* audits, as described in [7]. For our purposes, these types of audits are equivalent, since they both need to sample paper ballots at random, and can make use of the k-cut method proposed here. However, if one wishes to use k-cut sampling in a comparison audit, one would need to ensure that each paper ballot contains a printed ID number that could be used to locate the associated electronic CVR.

3 The *k*-Cut Method

The problem to be solved is:

> How can one select a single ballot (approximately) at random from a given stack of n ballots?

This section presents the "k-cut" sampling procedure for doing such sampling. The k-cut procedure does not need to know the size n of the stack, nor does it need any auxiliary random number generators or technology.

We assume that the collection of ballots to be sampled from is in the form of a stack. These may be ballots stored in a single box or envelope after scanning. One may think of the stack of ballots as being similar to a deck of cards. When the ballots are organized into *multiple* stacks, sampling is slightly more complex—see Section 7.

The basic operation for drawing a single ballot is called "k-cut and pick," or just "k-cut." This method does k cuts then draws the ballot at the top of the stack.

To make a single cut of a given stack of n paper ballots:

- Cut the stack into two parts: a "top" part and a "bottom" part.
- Switch the order of the parts, so what was the bottom part now sits above the top part. The relative order of the ballots within each part is preserved.

We let t denote the size of the top part. The size t of the top part should be chosen "fairly randomly" from the set $[n] = \{0, 1, 2, \ldots, n-1\}$[1]. In practice, cut sizes are probably not chosen so uniformly; so in this paper we study ways to compensate for non-uniformity. We can also view the cut operation as one that "rotates" the stack of ballots by t positions.

An example of a single cut. As a simple example, if the given stack has $n = 5$ ballots:

$$\boxed{\text{A B C D E}},$$

where ballot A is on top and ballot E is at the bottom, then a cut of size $t = 2$ separates the stack into a top part of size 2 and a bottom part of size 3:

$$\boxed{\text{A B}} \ \boxed{\text{C D E}}$$

whose order is then switched:

$$\boxed{\text{C D E}} \ \boxed{\text{A B}}\,.$$

Finally, the two parts are then placed together to form the final stack:

$$\boxed{\text{C D E A B}}.$$

having ballot C on top.

Iteration for k cuts. The k-cut procedure makes k successive cuts then picks the ballot at the top of the stack.

If we let t_i denote the size of the i-th cut, then the net rotation amount after k cuts is

$$r_k = t_1 + t_2 + \cdots + t_k \pmod{n}\,. \tag{1}$$

The ballot originally in position r_k (where the top ballot position is position 0) is now at the top of the stack. We show that even for small values of k (like $k = 6$) the distribution of r_k is close to $\mathcal{U}$.

[1] A cut of size n is excluded, as it is equivalent to a cut of size 0.

Drawing a sample of multiple ballots. To draw a sample of s ballots, our k-cut procedure repeats s times the operation of drawing without replacement a single ballot "at random." The s ballots so drawn form the desired sample.

Efficiency. Suppose a person can make six ("fairly random") cuts in approximately 15 seconds, and can count 2.5 ballots per second[2]. Then k-cut (with $k = 6$) is more efficient when the number of ballots that needs to be counted is 37.5 or more. Since batch sizes in audits are often large, k-cut has the potential to increase sampling speed.

For instance, assume that ballots are organized into boxes, each of which contains at least 500 ballots. Then, when the counting method is used, 85% of the time a ballot between ballot #38 and ballot #462 will be chosen. In such cases, one must count at least 38 ballots from the bottom or from the top to retrieve a single ballot. This implies that k-cut is more efficient 85% of the time.

As the number of ballots per box increases, the expected time taken by standard methods to retrieve a single ballot increases. With k-cut, the time it takes to select a ballot is *constant*, independent of the number of ballots in the box, assuming that each cut takes constant time.

Security We assume that the value of k is **fixed** in advance; you can not allow the cutter to stop cutting once a "ballot they like" is sitting on top.

4 (Non)-Uniformity of Single Ballot Selection

We begin by observing that if an auditor could perform "perfect" cuts, we would be done. That is, if the auditor could pick the size t of a cut in a perfectly uniform manner from $[n]$, then one cut would suffice to provide a perfectly uniform distribution of the ballot selected from the stack of size n. However, there is no *a priori* reason to believe that, even with sincere effort, an auditor could pick t in a perfectly uniform manner.

So, we start by studying the properties of the k-cut procedure for single-ballot selection, beginning with a study of the non-uniformity of selection for the case $k = 1$ and extending our analysis to multiple cuts.

4.1 Empirical Data for Single Cuts

This section presents our experimental data on single-cut sizes. We find that in practice, single cut sizes (that is, for $k = 1$) are "somewhat uniform." We then show that the approximation to uniformity improves dramatically as k increases.

We had two subjects, the authors. Each author had a stack of 150 sequentially numbered ballots to cut, provided by Marion County, Indiana. The authors made 1680 cuts in total. Figure 1 shows the observed cut size frequency distribution. The complete data tables are provided in the longer version of this paper [3].

[2] These assumptions are based on observations during the Indiana pilot audits.
[3] The longer version is available at https://arxiv.org/abs/1811.08811

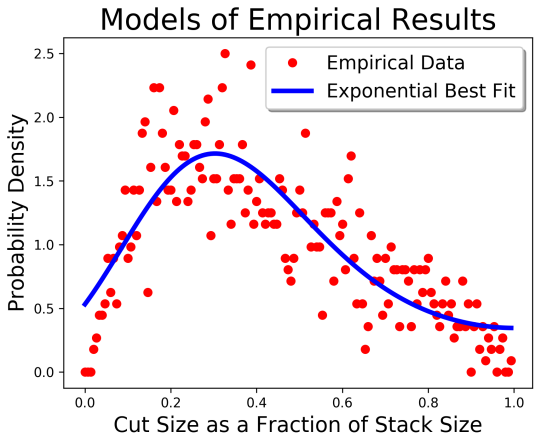

Fig. 1. Probability Density of empirical distribution of sizes of single cuts, using combined data from both authors, with 1680 cuts total. The model that best fit the empirical data was an exponential model, shown in blue. The extended paper provides more details about this and other models for our data.

If the cuts were truly random, we would expect a uniform distribution of the number of cuts observed as a function of cut size. In practice, the frequency of cuts was not evenly distributed; there were few or no very large or very small cuts, and smaller cuts were more common than larger cuts.

4.2 Making k successive cuts to select a single ballot

As noted, the distribution of cut sizes for a single cut is noticeably non-uniform. Our proposed k-cut procedure addresses this by iterating the single-cut operation k times, for some small fixed integer k.

We assume for now that cut sizes are distributed as in our experiments, as described in Figure 1, and that successive cuts are independent. Moreover, we assume that sampling is done with replacement, for simplicity.

We give computational results showing that as the number of cuts increases, the k-cut procedure selects ballots with a distribution that approaches the uniform distribution. We compare by computing the variation distance of the k-cut distribution from $\mathcal{U}$ for various k. We also computed ϵ, the maximum ratio of the probability of drawing any particular ballot under the empirical distribution, to the probability of drawing that ballot under the uniform distribution, minus one[4]. Our results are summarized in Table 1.

We can see that, after six cuts, we get a variation distance of about 7.19×10^{-4}, for the empirical distribution, which is often small enough to justify our recommendation that six cuts being "close enough" in practice, for any RLA.

[4] In Section 6.4, we discuss why this value of ϵ is relevant

k	Variation Distance	Max Ratio minus one
1	0.247	1.5
2	0.0669	0.206
3	0.0215	0.0687
4	0.0069	0.0224
5	0.00223	0.00699
6	**0.000719**	**0.00225**
7	0.000232	0.000729
8	7.49e-05	0.000235

Table 1. Convergence of k-cut to uniform with increasing k. Variation distance from uniform and ϵ-values for k cuts, as a function of k, for $n = 150$, where ϵ is one less than the maximum ratio of the probability of selecting a ballot under the assumed distribution to the probability of selecting that ballot under the uniform distribution.

4.3 Asymptotic Convergence to Uniform with k

As k increases, the distribution of cut sizes provably approaches the uniform distribution, under mild assumptions about the distribution of cut sizes for a single cut and the assumption of independence of successive cuts.

This claim is plausible, given the analysis of similar situations for continuous random variables. For example, Miller and Nigrini [9] have analyzed the summation of independent random variables modulo 1, and given necessary and sufficient conditions for this sum to converge to the uniform distribution.

For the discrete case, one can show that if once k is large enough that every ballot is selected by k-cut with some positive probability, then as k increases the distribution of cut sizes for k-cut approaches $\mathcal{U}$. Furthermore, the rate of convergence is exponential. The proof details are omitted here; however, the second claim uses Markov-chain arguments, where each rotation amount is a state, and the fact that the transition matrix is doubly stochastic.

5 Approximate Sampling

We have shown in the previous section that as we iterate our k-cut procedure, our distribution becomes quite close to the uniform distribution. However, our sampling still is not exactly uniform.

The literature on post-election audits generally assumes that sampling is perfect. One exception is the paper by Banuelos and Stark [2], which suggests dealing conservatively with the situation when one can not find a ballot in an audit, by treating the missing ballot as if it were a vote for the runner-up. Our proposed mitigation procedures are similar in flavor.

In practice, sampling for election audits is often done using software such as that by Stark [14] or Rivest [10]. Given a random seed and a number n of ballots to sample from, they can generate a pseudo-random sequence of integers from $[n]$, indexing into a list of ballot positions or ballot IDs. It is reasonable to

treat such cryptographic sampling methods as "indistinguishable from sampling uniformly," given the strength of the underlying cryptographic primitives.

However, in this paper we deal with sampling that is not perfect; the k-cut method with $k = 1$ is obviously non-uniform, and even with modest k values, as one might use in practice, there will be some small deviations from uniformity.

Thus, we address the following question:

> How can one effectively use an approximate sampling procedure in a post-election audit?

We let $\mathcal{G}$ denote the actual ("approximate") probability distribution over $[n]$ from the sampling method chosen for the audit. Our analyses assume that we have some bound on how close $\mathcal{G}$ is to $\mathcal{U}$, like variation distance. Furthermore, the quality of the approximation may be controllable, as it is with k-cut: one can improve the closeness to uniform by increasing k. We let $\mathcal{G}^s$ denote the distribution on s-tuples of ballots from $[n]$ chosen with replacement according to the distribution $\mathcal{G}$ for each draw.

6 Auditing Arbitrary Contests

This section proves a general result: for auditing an arbitrary contest, we show that *any* risk-limiting audit can be adapted to work with approximate sampling, if the approximate sampling is close enough to uniform. In particular, any RLA can work with the k-cut method, if k is large enough.

We show that if k is sufficiently large, the resulting distribution of k-cut sizes will be so close to uniform that any statistical procedure cannot efficiently distinguish between the two. That is, we want to choose k to guarantee that $\mathcal{U}$ and $\mathcal{G}$ are close enough, so that any statistical procedure behaves similarly on samples from each.

Previous work done by Baignères in [1] shows that, there is an optimal distinguisher between two finite probability distributions, which depends on the KL-Divergence between the two distributions.

We follow a similar model to this work, however, we develop a bound based on the variation distance between $\mathcal{U}$ and $\mathcal{G}$.

6.1 General Statistical Audit Model

We construct the following model, summarized in Figure 2.

We define δ to be the variation distance between $\mathcal{G}$ and $\mathcal{U}$. We can find an upper bound for δ empirically, as seen in Table 1. If $\mathcal{G}$ is the distribution of k-cut, then by increasing k we can make δ arbitrarily small.

The audit procedure requires a sample of some given size s, from $\mathcal{U}^s$ or $\mathcal{G}^s$. We assume that all audits behave deterministically. We do not assume that successive draws are independent, although we assume that each cut is independent.

Given the size s sample, the audit procedure can make a decision on whether to accept the reported contest result, escalate the audit, or declare an upset.

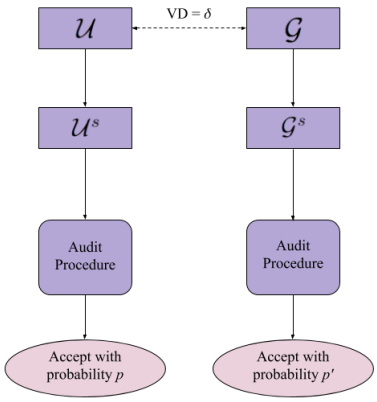

Fig. 2. Overview of uniform vs. approximate sampling effects, for any statistical auditing procedure. The audit procedure can be viewed as a distinguisher between the two underlying distributions. If it gives significantly different results for the two distributions, it can thereby distinguish between them. However, if p and p' are extremely close, then the audit cannot be used as a distinguisher.

6.2 Mitigation Strategy

When we use approximate sampling, instead of uniform, we need to ensure that the "risk-limiting" properties of the RLAs are maintained. In particular, as described in [7], an RLA with a risk limit of α guarantees that with probability at least $(1-\alpha)$ the audit will find and correct the reported outcome if it is incorrect. We want to maintain this property, while introducing approximate sampling.

Without loss of generality, we focus on the probability that the audit accepts the reported result, since it is the case where approximate sampling may affect the risk-limiting properties. We show that $\mathcal{G}$ and $\mathcal{U}$ are sufficiently close when k is large, that the difference between p and p', as seen in Figure 2, is small.

We show a simple mitigation procedure, for RLA plurality elections, to compensate for this non-uniformity, that we denote as **risk-limit adjustment.** For RLAs, we can simply decrease the risk limit α by $|p'-p|$ (or an upper bound on this) to account for the difference. This decrease in the risk limit can accommodate the risk that the audit behaves incorrectly due to approximate sampling.

6.3 How much adjustment is required?

We assume we have an auditing procedure $\mathbb{A}$, which accepts samples and outputs "accept" or "reject". We model approximate sampling with providing $\mathbb{A}$ samples from a distribution $\mathcal{G}$. For our analysis, we look at the empirical distribution of cuts in Table 2. For uniform sampling, we provide $\mathbb{A}$ samples from $\mathcal{U}$.

We would like to show that the probability that $\mathbb{A}$ accepts an outcome incorrectly, given samples from $\mathcal{G}$ is not much higher than the probability that $\mathbb{A}$

accepts an incorrect outcome, given samples from $\mathcal{U}$. We denote $\mathbb{B}$ as the set of ballots that we are sampling from.

Theorem 1. *Given a fixed sample size s and the variation distance δ, the maximum change in probability that $\mathbb{A}$ returns "accept" due to approximate sampling is at most*

$$\epsilon_1 + (1 + n\delta)^{s'} - 1,$$

where s' is the maximum number of "successes" seen in s Bernoulli trials, where each has a success probability of δ, with probability at least $1 - \epsilon_1$.

Proof. We define s as the number of ballots that we pull from the set of cast ballots, before deciding whether or not to accept the outcome of the election. Given a sample size s, based on our sampling technique, we draw s ballots, one at a time, from $\mathcal{G}$ or from $\mathcal{U}$.

We model drawing a ballot from $\mathcal{G}$ as first drawing a ballot from $\mathcal{U}$; however, with probability δ, we replace the ballot we draw from $\mathcal{U}$ with a new ballot from $\mathbb{B}$ following a distribution $\mathbb{F}$. We make no further assumptions about the distribution $\mathbb{F}$, which aligns with our definition of variation distance. When drawing from $\mathcal{G}$, for any ballot $b \in \mathbb{B}$, we have probability at most $\frac{1}{n} + \delta$ of drawing b.

When we sample sequentially, we get a length-s sequence of ballot IDs, S, for each of $\mathcal{G}$ and $\mathcal{U}$. Throughout this model, we assume that we sample with replacement, although similar bounds should hold for sampling without replacement, as well. We define X as the list of indices in the sequence S where both $\mathcal{G}$ and $\mathcal{U}$ draw the same ballot, in order. We define Z as the list of indices where $\mathcal{G}$ has "switched" a ballot after the initial draw. That is, for a fixed draw, $\mathcal{U}$ might produce the sample sequence [1, 5, 29]. Meanwhile, $\mathcal{G}$ might produce the sample sequence sequence [1, 5, 30]. For this example, $X = [0, 1]$ and $Z = [2]$.

We define the set of possible size-s samples as the set D. We choose s' such that for any given value ϵ_1, the probability that $|Z|$ is larger than s' is at most ϵ_1. Using this set up, we can calculate an upper bound on the probability that $\mathbb{A}$ returns "accept". In particular, given the empirical distribution, the probability that $\mathbb{A}$ returns "accept" for a deterministic auditing procedure becomes

$$\Pr[\mathbb{A}\ accepts \mid \mathcal{G}] = \sum_{S \in D} \Pr[\mathbb{A}\ accepts \mid S] * \Pr[draw\ S \mid \mathcal{G}]\,.$$

Now, we note that we can split up the probability that we can draw a specific sample S from the distribution $\mathcal{G}$. We know that with high probability, there are at most s' ballots being "switched". Thus,

$$\Pr[\mathbb{A}\ accepts \mid \mathcal{G}]$$

$$= \sum_{S \in D} \Pr[\mathbb{A}\, accepts \mid S] * \Pr[draw\, S \mid \mathcal{G}, S\ has\ \leq s'\ \text{"switched" ballots}] * \Pr[\text{S has } \leq s' \text{ "switched" ballots}]$$

$$+ \sum_{S \in D} \Pr[\mathbb{A}\, accepts \mid S] * \Pr[draw\, S \mid \mathcal{G}, S\ has\ > s'\ \text{"switched" ballots}] * \Pr[\text{S has } > s' \text{ "switched" ballots}]\,.$$

Now, we note that the second term is upper bounded by

$$\Pr[\text{any size-s sample has more than } s' \text{ switched ballots}]\,.$$

We define the probability that any size-s sample contains more than s' switched ballots as ϵ_1.

We note that, although the draws aren't independent, from the definition of variation distance, this is upper bounded by the probability that a binomial distribution, with s draws and δ probability of success.

Now, we can focus on bounding the first term. We know that

$$\Pr[\mathbb{A}\ accepts \mid \mathcal{G}, \text{any sample has at most } s' \text{ switched ballots}]$$

$$= \sum_{S \in D} \Pr[\mathbb{A}\ accepts \mid S] * \Pr[draw\ S \mid \mathcal{G}, S\ has\ \leq s'\ \text{“switched” ballots}]$$

For the uniform distribution, we know that the probability of accepting becomes

$$\Pr[\mathbb{A}\ accepts \mid \mathcal{U}] = \sum_{S \in D} \Pr[\mathbb{A}\ accepts \mid S] * \Pr[draw\ S \mid \mathcal{U}]\,.$$

Thus, we know that the change in probability becomes

$$\Pr[\mathbb{A}\ accepts \mid \mathcal{G}] - \Pr[\mathbb{A}\ accepts \mid \mathcal{U}]$$

$$\leq \epsilon_1 + \sum_{S \in D} \Pr[\mathbb{A}\ accepts \mid S](\Pr[draw\ S \mid \mathcal{G}, S\ has\ \leq s'\ \text{“switched” ballots}] - \Pr[draw\ S \mid \mathcal{U}])\,.$$

However, for any fixed sample S, we know that we can produce S from E in many possible ways. That is, we know that we have to draw at least $s - s'$ ballots that are from $\mathcal{U}$. Then, we have to draw the compatible s' ballots from $\mathcal{G}$. In general, we define the possible length $s - s'$ compatible shared list of indices as the set $\mathbb{X}$. That is, by conditioning on $\mathbb{X}$, we are now defining the exact indices in the sample tally where the uniform and empirical sampling can differ. We note that $|\mathbb{X}| = \binom{s}{s'}$ and each possible set happens with equal probability. Then, for any specific $x \in \mathbb{X}$, we can define z as the remaining indices, which are allowed to differ from uniform and approximate sampling. That is, if there are 3 ballots in the sample, and $x = [0, 1]$, then $z = [2]$.

We can now calculate the probability that we draw some specific size-s sample S, given the empirical distribution, and a fixed value of s'.

$$\Pr[draw\ S \mid \mathcal{G}] = \sum_{x \in \mathbb{X}} \Pr[draw\ x \mid \mathcal{U}] * \Pr[draw\ z \mid \mathcal{G}] * \Pr[\text{switched ballots are at indices in } z]$$

However, we know that for each ballot b in z, we draw ballot b with probability at most $\frac{1}{n} + \delta$. That is, for any ballot in x, we know that we draw it with uniform probability exactly. However, for a ballot b in z, we know that this a ballot that may have been "switched". In particular, with probability $\frac{1}{n}$, we draw the correct

ballot from $\mathcal{U}$. However, in addition to this, with probability δ, we replace it with a new ballot - we assume that we replace it with the correct ballot with probability 1. Thus, with probability at most $\frac{1}{n} + \delta$, we draw the correct ballot for this particular slot. Thus, we get

$$\begin{aligned}
&\Pr[\mathit{draw}\ S \mid \mathcal{G}] \\
&= \sum_{x \in \mathbb{X}} \Pr[\mathit{draw}\ x \mid \mathcal{U}] * \Pr[\mathit{draw}\ z \mid \mathcal{G}] * \Pr[\text{switched ballots are at indices in } z] \\
&\leq \sum_{x \in \mathbb{X}} \Pr[\mathit{draw}\ x \mid \mathcal{U}] * (\frac{1+n\delta}{n})^{s'} * \Pr[\text{switched ballots are at indices in } z] \\
&\leq (1+n\delta)^{s'} \sum_{x \in \mathbb{X}} \Pr[\mathit{draw}\ x \mid \mathcal{U}] * \Pr[\mathit{draw}\ z \mid \mathcal{U}] * \Pr[\text{switched ballots are at indices in } z]\,.
\end{aligned}$$

Now, we note that there are $\binom{s}{s'}$ possible sequences $x \in \mathbb{X}$, where the "switched" ballots could be. Each of these possible sequences occurs with equal probability, this becomes

$$\begin{aligned}
&\Pr[\mathit{draw}\ S \mid \mathcal{G}] \\
&\leq (1+n\delta)^{s'} \sum_{x \in \mathbb{X}} \Pr[\mathit{draw}\ x \mid \mathcal{U}] * \Pr[\mathit{draw}\ z \mid \mathcal{U}] * \Pr[\text{switched ballots are at indices in } z]\,. \\
&= (1+n\delta)^{s'} \sum_{x \in \mathbb{X}} \Pr[\mathit{draw}\ x \mid \mathcal{U}] * \Pr[\mathit{draw}\ z \mid \mathcal{U}] * \frac{1}{\binom{s}{s'}} \\
&= (1+n\delta)^{s'} \Pr[\mathit{draw}\ S \mid \mathcal{U}]\,.
\end{aligned}$$

Using this bound we can calculate our total change in acceptance probability as:

$$\Pr[\mathbb{A}\ \mathit{accepts} \mid \mathcal{G}] - \Pr[\mathbb{A}\ \mathit{accepts} \mid \mathcal{U}]$$

$$\begin{aligned}
&\leq \epsilon_1 + \sum_{S \in D} \Pr[\mathbb{A}\ \mathit{accepts} \mid S](\Pr[\mathit{draw}\ S \mid \mathcal{G}, S\ \mathit{has}\ \leq s'\ \text{“switched” ballots}] - \Pr[\mathit{draw}\ S \mid \mathcal{U}]) \\
&\leq \epsilon_1 + ((1+n\delta)^{s'} - 1) \sum_{S \in D} \Pr[\mathbb{A}\ \mathit{accepts} \mid S] \Pr[\mathit{draw}\ S \mid \mathcal{U}] \\
&\leq \epsilon_1 + (1+n\delta)^{s'} - 1\,,
\end{aligned}$$

which provides us the required bound.

6.4 Empirical Support

Our previous theorem gives us a total bound of our change in risk limit, which depends on our value of s' and δ. We note that, for each ballot b, we provide a general bound of a multiplicative factor increase of $(1 + n\delta)$, which is based off the variation distance of δ. However, we note that in practice, the exact

bound we are looking for depends on the multiplicative increase in probability of a single ballot being chosen. That is, we can calculate the max increase in multiplicative ratio for a single ballot, compared to the uniform distribution. Thus, if a ballot is chosen with probability at most $\frac{(1+\epsilon_2)}{n}$, then our bound on the change in probability becomes

$$\epsilon_1 + (1+\epsilon_2)^{s'} - 1.$$

The values of ϵ_2 are recorded, for varying number of cuts in Table 1.

We can calculate the maximum change in probability for a varying number of cuts using this bound. Here, we analyze the case of 6 cuts. To get a bound on s', we can model how often we switch ballots. In particular, this follows a binomial distribution, with s independent trials, where each trial has a δ_6 probability of success. Using the binomial survival function, we see at most 4 "switched ballots" in 1,000 draws, with probability (1- 8.78×10^{-4}). From our previous argument, we know that our change in acceptance probability is at most $(1+\epsilon_2)^4 - 1$. Using our value of ϵ_2 for $k = 6$, this causes a change in probability of at most 0.0090.

Thus, the maximum possible change in probability of incorrectly accepting this outcome is $0.0090 + 8.78 \times 10^{-4}$, which is approximately 9.88×10^{-3}. We can compensate for this by adjusting our risk limit by less than 1%.

7 Multi-stack Sampling

Our discussion so far presumes that all cast paper ballots constitute a single "stack," and suggest using our proposed k-cut procedure is used to sample ballots from that stack. In practice, however, stacks have limited size, since large stacks are physically awkward to deal with. The collection of cast paper ballots is therefore often arranged into multiple stacks of some limited size.

The *ballot manifest* describes this arrangement of ballots into stacks, giving the number of such stacks and the number of ballots contained in each one. We assume that the ballot manifest is accurate. A tool like Stark's Tools for Risk-Limiting Audits [5] takes the ballot manifest (together with a random seed and the desired sample size) as input and produces a sampling plan.

A sampling plan describes exactly which ballots to pick from which stacks. That is, the sampling plan consists of a sequence of pairs, each of the form: (stack-number, ballot-id), where ballot-id may be either an id imprinted on the ballot or the position of the ballot in the stack (if imprinted was not done).

Modifying the sampling procedure to use k-cut is straightforward. We ignore the ballot-ids, and note only how many ballots are to be sampled from each stack. That number of ballots are then selected using k-cut rather than using the provided ballot-ids. For example, if the sampling plan says that 2 ballots are to be drawn from stack 5, then we ignore the ballot-ids for those specific ballots, and return 2 ballots drawn approximately uniformly at random using k-cut.

Thus, the fact that cast paper ballots may be arranged into multiple stacks (or boxes) does not affect the usability of k-cut for performing audits.

[5] https://www.stat.berkeley.edu/ stark/Vote/auditTools.htm

8 Approximate Sampling in Practice

The major question when using the approximate sampling procedure is how to choose k. Choosing a small value of k makes the overall auditing procedure more efficient, since you save more time in each sample you choose. However, it requires more risk limit adjustment.

The risk limit mitigation procedure requires knowledge of the maximum sample size, which we denote as s^*, beforehand. We assume that the auditors have a reasonable procedure for estimating s^* for a given contest. One procedure to estimate s^* is to draw an initial sample, s, using uniform random sampling. Then, we can use a statistical procedure to approximate how many additional ballots we would need to finish the audit, assuming the rest of the ballots in the pool are similar to the sample. Possible statistical procedures include replicating the votes on the ballots, or using sample size estimates defined in [8].

Let us assume that we use one of these techniques and calculate that the audit is complete after an extension of size d. To be safe, we can assume that at most $3d$ additional samples will be needed. Thus, our final bound on s^* would be $s + 3d$. Given this upper bound, we can perform our mitigation procedures, assuming that we are drawing a sample of size s^*. Ballots after the first s^* ballots in our sample should be sampled uniformly at random.

9 Discussion and Open Problems

We would like to do more experimentation on the variation between individuals on their cut-size distributions. The current empirical results in this paper are based off of the cut distributions of just the two authors in the paper. We would like to test a larger group of people to better understand a variety of empirical distributions. After investigating this, we would like to develop "best practices" for using the k-cut procedure. That is, we'd like to develop a set of techniques that auditors can use to produce nearly-uniform single-cut-size distributions, which will make k-cut more efficient.

We would also like to run some experiments to test our assumptions for k-cut, in practice. For instance, we would like to test whether each cut is truly made independently.

In the longer version of the paper, we provide the full details of our empirical data, for full reproducibility. We also discuss possible models for our empirical data and the convergence rates of our models.

10 Conclusions

We have presented an approximate sampling procedure, k-cut, for use in post-election audits. We expect the use of k-cut will save time since it eliminates the need to count many ballots in a stack to find the desired one.

We showed that even for small values of k, our procedure provides a sample that is close to being chosen uniformly at random. We designed a simple mitigation procedure for RLAs that accounts for any remnant non-uniformity, by adjusting the risk limit. Finally, we provided a recommendation of $k = 6$ cuts to use in practice, for sample sizes up to 1,000 ballots, based on our empirical data, with a 1% risk limit adjustment.

An earlier version of k-cut was used in pilot audits in Marion County, Indiana to increase audit efficiency. This paper provides theoretical justification for this technique, which is also scheduled to be used in Michigan in December 2018.

References

1. Baignères, T., Vaudenay, S.: The complexity of distinguishing distributions (2008), results also in Baigneères' PhD thesis.
2. Banuelos, J.H., Stark, P.B.: Limiting risk by turning manifest phantoms into evil zombies. https://arxiv.org/abs/1207.3413 (2012)
3. Bretschneider, J., Flaherty, S., Goodman, S., Halvorson, M., Johnston, R., Lindeman, M., Rivest, R., Smith, P., Stark, P.: Risk-limiting post-election audits: Why and how? (Oct 2012), (ver. 1.1) http://people.csail.mit.edu/rivest/pubs.html#RLAWG12
4. Goggin, S.N., Byrne, M.D., Gilbert, J.E.: Post-election auditing effects of procedure and ballot type on manual counting accuracy, efficiency, and auditor satisfaction and confidence. Election Law Journal (2012)
5. Johnson, K.: Election verification by statistical audit of voter-verified paper ballots. http://ssrn.com/abstract=640943 (Oct 31 2004)
6. Lindeman, M., Halvorseon, M., Smith, P., Garland, L., Addona, V., McCrea, D.: Principle and best practices for post-election audits. www.electionaudits.org/files/best%20practices%20final_0.pdf (2008)
7. Lindeman, M., Stark, P.B.: A gentle introduction to risk-limiting audits. IEEE Security and Privacy **10**, 42–49 (2012)
8. Lindeman, M., Stark, P.B., Yates, V.S.: BRAVO: Ballot-polling risk-limiting audits to verify outcomes. In: Halderman, A., Pereira, O. (eds.) Proceedings 2012 EVT/WOTE Conference (2012)
9. Miller, S.J., Nigrini, M.J.: The modulo 1 Central Limit Theorem and Benford's law for products. https://arxiv.org/abs/math/0607686 (2007)
10. Rivest, R.L.: Reference implementation code for pseudo-random sampler. http://people.csail.mit.edu/rivest/sampler.py (2011)
11. Rivest, R.L.: Bayesian tabulation audits: Explained and extended. https://arxiv.org/abs/1801.00528 (January 1, 2018)
12. Rivest, R.L., Shen, E.: A Bayesian method for auditing elections. In: Halderman, J.A., Pereira, O. (eds.) Proceedings 2012 EVT/WOTE Conference (2012), https://www.usenix.org/system/files/conference/evtwote12/rivest_bayes_rev_073112.pdf, https://www.usenix.org/conference/evtwote12/workshop-program/presentation/rivest
13. Stark, P.B.: Papers, talks, video, legislation, software, and other documents on voting and election auditing. https://www.stat.berkeley.edu/ stark/Vote/index.htm
14. Stark, P.B.: Tools for ballot-polling risk-limiting election audits. https://www.stat.berkeley.edu/ stark/Vote/ballotPollTools.htm (2017)