On the Randomized Improvement of Markov's Inequality

Christian Covington
Dominic DiSanto

Notes/Last fixes

- Write conclusion (I can do Wednesday at airport, leaving for now to be a bit lazy)

  Christian: no prob, I can also take a look at it

1 Background

In their 2023 paper *Randomized and exchangeable improvements of Markov's, Chebyshev's and Chernoff's inequalities*, Ramdas and Manole provide randomized improvements of Markov's (and related) inequalities [RM23]. The authors provide simple extensions of these tail-bound inequalities with the introduction of a standard uniform random variable, leading to stronger statements from which the original (hereafter referred to as the "naïve" versions) inequalities can be recovered. We begin briefly with an overview of the specific bounds proposed in [RM23] in 2 as an introduction to these results.

In Section 3, we simulate data and apply these randomized inequalities to examine the behavior and relative improvement of uniform random bounds over the use of naïve Markov's. Inspired by the typical Bonferroni multiple comparison correction procedure and the authors' comment on "U-hacking" , we also explored an initial idea to use a scaled minimum order statistic of a random sample from the standard uniform, which we compare to the performance of naïve and randomized inequalities.

From the results of our work in Section 3, and perhaps naturally, one may also question the "optimality" of the proposed randomization. Would other random variables satisfy Definition 4.1 and improve on the use of the standard uniform? We consider this question in Section 4, where we more explicitly state what we feel to be a reasonable "optimality" consideration and demonstrate that the standard uniform satisfies this optimality condition. The resulting Theorem 4.4 is the contribution of most interest, where we identify that in fact the standard uniform random variable is an optimal choice of randomizer for the familiy of randomized inequalities which holds with no additional restrictions on X, a.

2 Randomized Bounds

The idea advanced in [RM23] is that standard concentration inequalities (e.g. Markov, Chebyshev, Chernoff) are improvable by "randomized" generalizations. For example, the standard Markov's Inequality is as follows:

Theorem 2.1 (Markov's Inequality)**.** *Let X be a non-negative random variable. Then for any $a > 0$ we know that*

$$\mathbb{P}\left(X \geqslant a\right) \leqslant \frac{\mathbb{E}\,X}{a}.$$

The Uniformly-Randomized Markov's Inequality (UMI) from [RM23] is the following similar, but stronger, statement:

Theorem 2.2 (Uniformly-Randomized Markov's Inequality)**.** *Let X be a non-negative random variable and U a random variable, independent of X, which is stochastically larger than* Unif$(0, 1)$. Then for any $a > 0$ we know that

$$\mathbb{P}\left(X \geqslant aU\right) \leqslant \frac{\mathbb{E}\,X}{a}.$$

with a simple, one-line proof which we present here from the original work [RM23]:

$$\mathbb{P}(X \geqslant aU) = \mathbb{P}(X/a \geqslant U) = \mathbb{E}\left[\mathbb{P}(X/a \geqslant U|X)\right] = \mathbb{E}\min\{1, X/a\} \leqslant \mathbb{E}(X)/a$$

As previously mentioned, similar to the extensions of naïve Markov, we can extend the randomized bound to other useful inequalities or tail bounds. We omit the proofs but re-state two examples of these extensions from [RM23], as well as their proposed general definition:

Theorem 2.3 (Uniformly-Randomized Chebyshev's Inequality)**.** *For iid samples of random variable X with variance $V(X) = \sigma^2$, $U \sim$Unif$(0,1)$ independent of X, and $a > 0$:*

$$\mathbb{P}\left(|\overline{X} - \mathbb{E}\,X| \geqslant a\sigma\sqrt{\frac{U}{n}}\right) \leqslant \frac{1}{a^2}$$

This bound also gives rise to a uniformly-randomized $(1-\alpha)$ confidence interval:

$$\overline{X} \pm \frac{\sigma U}{\sqrt{n\alpha}}$$

Theorem 2.4 (Uniformly-Randomized Hoeffding's Inequality)**.** *For iid samples of sub-Gaussian random variable X with variance proxy σ and $U \sim$Unif$(0,1)$ independent of X, then the following tail bound holds $\forall\ \alpha \in (0,1)$:*

$$\mathbb{P}\left(\sigma\sqrt{\frac{2\log(1/\alpha)}{n}} + \frac{\log(U)}{\sqrt{2n\log(1/\alpha)}}\right) \leqslant \alpha$$

which similarly gives the form for a $(1-\alpha)$ confidence interval for the sample mean of $\sigma-$sub-Gaussian random variables

$$\overline{X} \pm \left(\sigma\sqrt{\frac{2\log(2/\alpha)}{n}} + \frac{\log(U)}{\sqrt{2n\log(2/\alpha)}}\right)$$

Lastly, having observed the structure of these example, randomized tail bounds, we present the generalized definition from the authors' Proposition C.1:

Theorem 2.5 (General Randomized Inequality)**.** *For random variable X in $\mathcal{X} \subseteq \mathbb{R}$ and interval $I \subseteq \mathbb{R}^{\geqslant 0}$, and two nondecreasing, Borel-measurable functions $f : \mathcal{X} \mapsto I, g : I \mapsto \mathcal{X}$ such that $f(g(z)) \geqslant z, \forall z \in I$, then the following holds for any $U \sim$ Unif$(0,1)$ independent of X and any $x > 0$:*

$$\mathbb{P}(X \geqslant g(Uf(x))) \leqslant \frac{\mathbb{E}\,f(X)}{f(x)}$$

which encompasses the previous, randomized statements of Markov's, Chebyshev's, and Hoeffding's respective inequalities. Note that the above theorems hold for U $\sim$Unif$(0,1)$ as well as any random variable that is stochastically larger than the standard uniform (and independent of X).

3 Empirical Examination via Simulations

We have so far observed our ability to construct randomized inequalities tighter than their naïve representations and the ability to construct (presumably) tighter confidence intervals as a result. As an initial exploration of the method, it is of interest to observe the seeming gain (i.e. tightening of confidence interval width) by using these randomized extensions.

Additionally, the authors acknowledge the risk of U-hacking, that is taking random samples of uniform (or stochastically larger) random variables and selecting the smallest or most convenient value for a desired result. This is clearly an abuse of the proposed method and would not guarantee the desired coverage of generated intervals. However it does raise an interesting question: can the result of randomized bounds be improved by the inclusion of a larger random sample of U with some correction that results in the desired behavior (e.g. true $(1-\alpha)$ coverage in confidence intervals).

In an attempt to explore this hypothesis, we applied a naïve Bonferroni-style correction to confidence intervals constructed using the randomized Chebyshev inequality. We compare a naïve Chebyshev inequality (no random element included), the "typical" Chebyshev randomized inequality (single sample), and a series of "adjusted" inequalities. Letting $U_{(1)}$ represents the minimum order statistic from some *iid* sample of k standard uniform random variables, we compare the following methods of confidence interval construction:

$$\text{Naïve CI}_{\text{Naïve}} = \overline{X}_n \pm \frac{\sigma}{\sqrt{\alpha n}}$$

$$\text{“Typical” Random Chebyshev CI} = \overline{X}_n \pm \frac{\sigma\sqrt{U}}{\sqrt{\alpha n}}$$

$$\text{“Adjusted” Random Chebyshev CI} = \overline{X}_n \pm \frac{\sigma\sqrt{kU_{(1)}}}{\sqrt{\alpha n}}$$

For a fixed sample $n = 1000$, we calculated $\overline{X}_n = n^{-1}\sum_i X_i$ and constructed confidence intervals as outlined above, with our adjusted intervals calculated for $k = \{1, 20, 100, 1000\}$. We additionally calculate a Hoeffding-derived $(1-\alpha)$ interval as described in Section 2. This was repeated for $B = 2000$ iterations. For comparison we also include the 95% interval for the Gaussian mean as $\overline{X} \pm z_{1-\alpha/2}/\sqrt{n}$, which can be seen as using our *a priori* knowledge of the Gaussianity of X or the known asymptotic distribution of the sample mean, invoking the Central Limit Theorem if interested in extending to non-Guassian variables.

Figure 1 first plots all observed confidence intervals by iteration, with two vertically-aligned dots for each iteration value representing the confidence interval. The Gaussian and Naïve panels are of constant width, i.e. include no randomization. We see the randomized plots include a number of intervals much tighter than the naïve (and even true-Gaussian) intervals permit. We observe that the typical ($k = 1$) randomized interval is universally tighter than the naïve Chebyshev interval, while our adjusted $k = 20, 100, 1000$ intervals are out performed by the $k = 1$ typical interval and in fact are not necessarily tighter than the naïve interval. We do observe that the Hoeffding randomized bound is the tightest among all non-exact Gaussian intervals, which may be expected as the use of Hoeffding in this setting utilizes our knowledge of subGaussianity (in fact Gaussianity) of X.

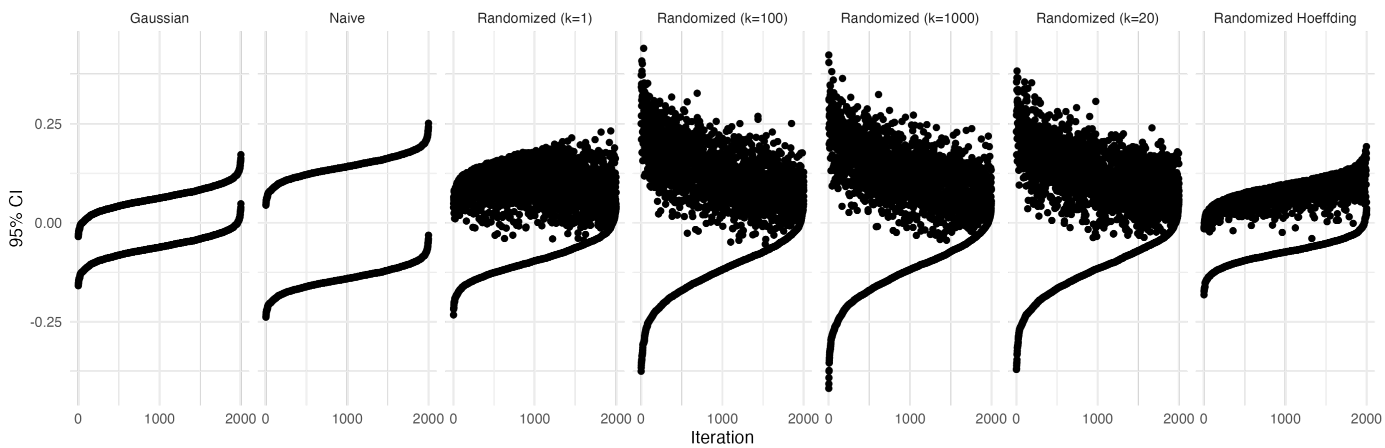

Figure 1: Comparison of 95% Confidence Intervals by Tail-Bound Randomization for Gaussian Mean (n=1000)

Figure 2 summarizes the interval width over our $B = 2000$ iterations. The Naïve and Gaussian plots are constant values (fixed-width confidence intervals). The four central plots are randomized-Chebyshev generated confidence intervals, where $k = 1$ is the typical interval and larger value sof k our adjusted intervals. We observe that our adjusted intervals do not generally improve upon the single-observation or typical randomized Chebyshev interval.

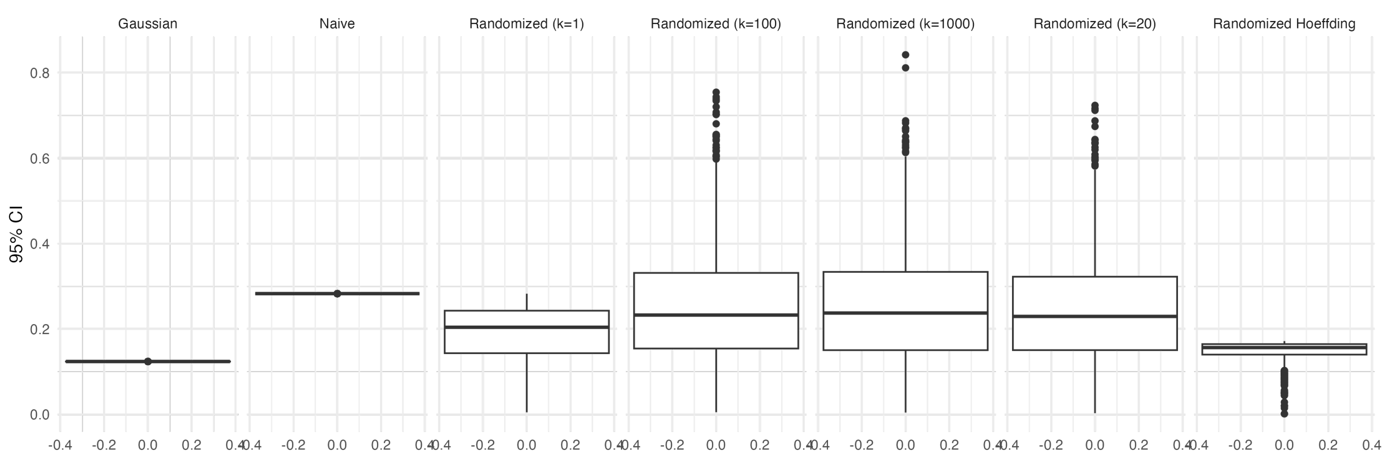

Figure 2: Comparison of Interval Width by Tail-Bound Randomization for Gaussian Mean (n=1000)

We conclude by examining the empirical coverage probability of our methods, i.e. $\sum_{b=1}^{B} \mathbf{1}(0 \in \mathrm{CI}_b)$. As expected we see the random-Chebyshev-generated intervals are less conservative than the naïve Chebyshev generated intervals, while coverage is largely similar (within Monte Carlo error) across our adjusted-randomized-Chebyshev generated intervals.

Type	Coverage Probability
Gaussian	0.95
Naïve	1.00
Randomized Markov's (k=1)	0.95
Randomized Markov's (k=100)	0.95
Randomized Markov's (k=1000)	0.96
Randomized Markov's (k=20)	0.95
Randomized Hoeffding	0.97

Table 1: Table of Empirical Confidence Interval Coverage Probability over B=2000 Iterations

4 Considering the optimality of the Uniform

Although our previous analysis was unfruitful in improving upon the randomized bounds (or downstream confidence intervals generated by said bounds) proposed by [RM23], we then began to consider that the choice of U ∼Unif$(0,1)$ could in fact be optimal in some sense. To this end, it is first necessary to define the general property of being a valid randomizer and to articulate the desired behavior (and thus optimality) of randomized bounds as defined. We did so by focusing on the randomized Markov inequality, under the assumption that optimality for the randomized Markov bound carries forward into the Markov-derived inequalities (e.g. Chebyshev, Hoeffding).

We begin by defining a valid MI randomizer and noting a useful definition and language for stochastic ordering that we will use:

Definition 4.1 (Valid Markov's Inequality Randomizer)**.** *A random variable $R : \Omega \to \mathbb{R}^{\geqslant 0}$ is a valid Markov's inequality randomizer if*

$$\forall X : \Omega \to \mathbb{R}^{\geqslant 0}, \forall a \geqslant 0 : \mathbb{P}\left(X \geqslant aR\right) \leqslant \frac{\mathbb{E}\, X}{a}.$$

Definition 4.2 (Stochastic Order)**.** *Given random variables X, Y, we say that Y is stochastically larger than X, denoted $X \preceq Y$, if*

$$\forall a \in \mathbb{R} : \mathbb{P}(Y \leqslant a) \leqslant \mathbb{P}(X \leqslant a).$$

If $X \preceq Y$, then we also say X is stochastically smaller than Y.

We can make this ordering strict, denoted $X \prec Y$, if $X \preceq Y$ but also there exists at least one $a \in \mathbb{R}$ such that $\mathbb{P}(Y \leqslant a) < \mathbb{P}(X \leqslant a)$.

We must now articulate more carefully what "optimality" is desired for a randomized version of the Markov Inequality and whether or not the Unif$(0,1)$ is optimal by any/all plausible definitions.

Definition 4.3 (Optimal Markov's Inequality Randomizer)**.** *A valid Markov's inequality randomizer R is the optimal such randomizer if R for any other valid randomizer R':*

$$\forall X : \Omega \mapsto \mathbb{R}^{\geqslant 0}, a \geqslant 0 : \mathbb{P}(X \geqslant aR) \leqslant \mathbb{P}(X \geqslant aR')$$

This definition gives us that the optimal randomizer is the tighest possible construction of a randomized Markov's inequality that holds for all positive random variable X and positive constant a. We know from [RM23] that U ∼Unif$(0,1)$ and anything stochastically larger than it are valid MI randomizers. Clearly however any variable stochastically larger than the standard uniform is sub-optimal, as the resulting randomized bound is improved by using a standard uniform randomizer.

Thus any possible improvement in a randomized Markov must either be stochastically smaller than Unif$(0,1)$ or not stochastically dominated/dominating a Unif$(0,1)$ random variable. Below we prove that these classes of distributions **cannot** be valid Markov's inequality randomizers, which hinge upon an argument that any variable R that is not stochastically larger than U must be stochastically smaller on some (if not all) subset of the common support $\mathcal{X} \cap \mathcal{R}$ (or more simply some subset of $(0,1)$):

Theorem 4.4. *Let $U \sim$Unif$(0,1)$. If $R \not\succeq U$, R cannot be a valid MI randomizer.*

Proof. Because $R \not\succeq U$, there exists some $y \in \mathbb{R}^{\geqslant 0}$ such that $\mathbb{P}(U \leqslant y) < \mathbb{P}(R \leqslant y)$. Because $\forall z \geqslant 1 : \mathbb{P}(U \leqslant z) = 1$, we know in fact that this $y \in (0,1)$.

Let $a = y^{-1}$ and define $X \sim \text{Bern}(p)$ for some $p \in (0,1)$. Then we have the following randomized bound:

$$
\begin{aligned}
\mathbb{P}\left(X \geqslant aR\right) &= \mathbb{P}\left(R \leqslant \frac{X}{a}\right) \\
&= \mathbb{E}\,\mathbb{P}\left(R \leqslant \frac{X}{a}|X\right) \\
&= p\,\mathbb{P}\left(R \leqslant \frac{X}{a}|X=1\right) + (1-p)\,\mathbb{P}\left(R \leqslant \frac{X}{a}|X=0\right) \\
&> p\,\mathbb{P}\left(U \leqslant \frac{X}{a}|X=1\right) + (1-p)\,\mathbb{P}\left(U \leqslant \frac{X}{a}|X=0\right) \quad \left(\mathbb{P}(U \leqslant y) < \mathbb{P}(R \leqslant y) \text{ and } y = a^{-1}\right) \\
&= \frac{p}{a} \\
&= \frac{\mathbb{E}\,X}{a}.
\end{aligned}
$$

Thus, we found a pair of X, a such that the Markov Inequality is violated when randomizing with R. Because the inequality must hold for all X, a, we know R cannot be a valid MI randomizer. □

Corollary 4.5 (Optimality of U as Markov's Inequality Randomizer)**.** *Let U* $\sim$Unif$(0,1)$, then U is the optimal Markov's Inequality Randomizer, in the sense of Definition 4.3.

Proof. We know via contraposition of Theorem 4.4 that, for any valid randomizer R, $U \preceq R$.

Let $X : \Omega \mapsto \mathbb{R}^{\geqslant 0}$ and $a^{-1} \geqslant 0$ be arbitrary. Then we have

$$
\begin{aligned}
U \preceq R &\iff \mathbb{P}\left(R \leqslant a^{-1}\right) \leqslant \mathbb{P}\left(U \leqslant a^{-1}\right) \\
&\implies \mathbb{P}\left(\frac{R}{x} \leqslant a^{-1}\right) \leqslant \mathbb{P}\left(\frac{U}{x} \leqslant a^{-1}\right) \quad (\text{for all } x \geqslant 0) \\
&\implies \mathbb{P}\left(\frac{R}{X} \leqslant a^{-1}\right) \leqslant \mathbb{P}\left(\frac{U}{X} \leqslant a^{-1}\right) \\
&\iff \mathbb{P}\left(X \geqslant aR\right) \leqslant \mathbb{P}\left(X \geqslant aU\right),
\end{aligned}
$$

which exactly matches Definition 4.3. □

5 Conclusion

Appendix

A References

[RM23] Aaditya Ramdas and Tudor Manole. Randomized and exchangeable improvements of markov's, chebyshev's and chernoff's inequalities. *arXiv preprint arXiv:2304.02611*, 2023.