# SENIOR CAPSTONE PROJECT

## Performance of the Chen-Stein Method on Sums of Bernoulli Random Variables

Nicole Dong

Department of Computer Science

Department of Applied Mathematics

University of Colorado - Boulder

Defended on April 28 2022

Committee Members

Thesis Advisor: Manuel E. Lladser, Department of Applied Mathematics (APPM)

Joshua A. Grochow, Department of Computer Science (CSCI)

Bo Waggoner, Department of Computer Science (CSCI)

**Abstract**

This project studies an extension of the Chen-Stein method based on the negative binomial distribution. We prove some lemmas used implicitly in the literature to estimate the total variation distance between a random variable of the form $S := \sum_{i=1}^{n} X_i$, where $X_1, \ldots, X_n$ are possibly dependent binary random variables, and a random variable $W$ with a suitable negative binomial distribution. We also examine the accuracy of this approximation considering different models of dependence between $X_1, \ldots, X_n$.

# Table of Contents

# 1 Introduction

The Chen-Stein method is a method developed by Stein [7] and Chen [1] to obtain bounds on the error of approximating the distribution of a random variable by another, usually with a simpler distribution. In the context of this project, the error of approximating one probability distribution by another one is measured using the *total variation distance*. For probability distributions $\mu$ and $\nu$ over $\mathbb{N} := \{0, 1, 2, \ldots\}$—the setting of this project—this is defined as

$$\|\mu - \nu\| := \sup_{A \subset \mathbb{N}} |\mu(A) - \nu(A)| = \frac{1}{2} \sum_{n=0}^{\infty} |\mu(n) - \nu(n)|. \tag{1}$$

The later identity holds because the supremum is achieved at the sets $\{n \in \mathbb{N} \text{ such that } \mu(n) \geq \nu(n)\}$ and $\{n \in \mathbb{N} \text{ such that } \mu(n) < \nu(n)\}$.

The Chen-Stein method has numerous theoretical applications ranging from the famous Birthday Problem to the total number of cycles in a random graph [5]. On the more practical side, a recurring problem in molecular biology requires determining whether or not two different DNA strands are similar just by chance [6]. In this case, each strand is described by its sequence of DNA bases (A, C, G, and T). Given an integer $\ell > 0$, a standard way to score how similar are two strands is to compute the greatest number of matches between any two sub-strings of length $\ell$. The Chen-Stein method is useful for approximating the distribution of this score.

Traditionally, the Chen-Stein method has been used to approximate the distribution of a random variable of the form $\sum_{i=1}^{n} X_i$, where $X_1, \ldots, X_n$ are binary. In the special case that these are *independent and identically distributed* (i.i.d.), the Chen-Stein method provides an upper-bound to the total variation distance between the binomial and Poisson distribution—see equation (2) for details. The method can also be applied to approximate the distribution of $\sum_{i=1}^{n} X_i$ by a Poisson distribution under various hypotheses that implicitly assume a low level of dependence between the Bernoulli random variables in the summation.

A discrete random variable (or its probability distribution) is called *under-dispersed* when its expected value is larger than its variance. If instead its expected value is smaller than its variance, it is called *over-dispersed*. Dispersion can guide the selection of one probability distribution over another to approximate the distribution of a highly complex discrete random variable.

The Poisson distribution is neither under- nor over-dispersed because its expected value and variance are equal. This feature is undesirable in some applications when the sample mean of a random quantity is much smaller than its sample variance. An example of this is the number of "reads" (i.e. relatively short DNA sequences observed when sequencing DNA) that map to a given genomic window [3, 2]. This has motivated practitioners to use the negative binomial distribution—which is over-dispersed—to approximate the probability distribution of

$$S := \sum_{i=1}^{n} X_i,$$

when $X_1, \ldots, X_n$ are possibly dependent binary random variables.

In this project, we study an extension of the Chen-Stein method based on the negative binomial distribution. In Section 2, we show some lemmas used implicitly in the literature to find an upper-bound for the total variation distance between $S$ and a random variable $W$ with a suitable negative binomial distribution, while referencing the techniques used in the Chen-Stein method based on a Poisson distribution approximation. In Section 3, we study different sums of Bernoulli random variables with various levels of dependence. We also examine their theoretical total variation distance according to [8] as well as their true total variation distance according to experimental data.

We finish the Introduction revising some relevant distributions.

## 1.1 Bernoulli and Binomial Distributions

In what follows, $X \sim \text{Bernoulli}(p)$ denotes a random variable that follows a *Bernoulli distribution* with parameter $p \in [0, 1]$. In particular, its *probability mass function* (p.m.f.) is given by:

$$\mathbb{P}(X = 0) := 1 - p;$$
$$\mathbb{P}(X = 1) := p.$$

The event $X = 1$ is usually described as a *success*, while event $X = 0$ is usually described as a *failure*.

On the other hand, $Y \sim \text{Binomial}(n, p)$ denotes a random variable with a *binomial distribution* with parameters $n > 0$ an integer and $p \in [0, 1]$. The p.m.f. of $Y$ is

$$\mathbb{P}(Y = y) = \binom{n}{y} p^y (1 - p)^{n-y}, \text{ for integers } y = 0, \ldots, n.$$

The expected value and variance of $Y$ are

$$\begin{aligned}\mathbb{E}(Y) &= n\,p;\\ \mathbb{V}(Y) &= n\,p\,(1-p).\end{aligned}$$

Since $\mathbb{V}(Y) = \mathbb{E}(Y)\cdot(1-p)$, $\mathbb{E}(Y) > \mathbb{V}(Y)$, the binomial distribution is under-dispersed.

Note that a Bernoulli random variable corresponds to a binomial random variable with parameter $n = 1$. In particular, if $X \sim$ Bernoulli$(p)$ then $\mathbb{E}(X) = p$ and $\mathbb{V}(X) = p\,(1-p)$.

The binomial distribution can be interpreted as the total number of successes in a sequence of $n$ independent Bernoulli trials, each with probability $p$ of success. Accordingly, if $X_1, \ldots, X_n$ are i.i.d. Bernoulli random variables with the same parameter $p$ then

$$\sum_{i=1}^{n} X_i \sim \text{Binomial}(n, p).$$

## 1.2 Poisson Distribution

We write $Z \sim$ Poisson$(\lambda)$ to denote a random variable that follows a *Poisson distribution* with parameter $\lambda > 0$. The p.m.f. of $Z$ is

$$\mathbb{P}(Z = z) = \frac{\lambda^z e^{-\lambda}}{z!}, \text{ for integers } z = 0, 1, \ldots$$

The expected value and variance of $Z$ are

$$\mathbb{E}(Z) = \mathbb{V}(Z) = \lambda.$$

Let $X_1, \ldots, X_n$ be independent but *not necessarily* equally distributed Bernoulli random variables, say $X_i \sim$ Bernoulli$(p_i)$ with $p_i \in [0, 1]$. Define

$$S := \sum_{i=1}^{n} X_i.$$

The distribution of $S$ may be approximated by that of a Poisson random variable with parameter $\lambda := \mathbb{E}(S) = \sum_{i=1}^{n} p_i$. More precisely, if

$$Z \sim \text{Poisson}\left(\sum_{i=1}^{n} p_i\right)$$

then

$$\left\| \sum_{i=1}^{n} X_i - Z \right\| \le \sum_{i=1}^{n} p_i^2.$$

In particular, for example, if $p_1 = \cdots = p_n = p$ then $\lambda = np$ and the above inequality translates into

$$\max_{A \subset \{0,1,\ldots,n\}} \left| \sum_{k \in A} \binom{n}{k} p^k\,(1-p)^{n-k} - \sum_{k \in A} \frac{(np)^k e^{-np}}{k!} \right| \le np^2. \tag{2}$$

## 1.3 Negative Binomial Distribution

We write $W \sim$ NegBin$(r, p)$ to denote a random variable that follows a *negative binomial distribution* with parameters $r > 0$ an integer and $p \in [0, 1]$. Its p.m.f. is as follows:

$$\mathbb{P}(W = w) = \binom{w + r - 1}{w} (1-p)^r\, p^w, \text{ for integers } w \ge 0.$$

The expected value and variance of $W$ are given by:

$$\mathbb{E}(W) = \frac{rp}{1-p};$$
$$\mathbb{V}(W) = \frac{rp}{(1-p)^2}.$$

Generally, we use $q$ to denote $(1-p)$. We note that there are different formulations of the negative binomial distribution. The one we have adopted models the total number of successes before $r$ failures occur in a sequence of i.i.d. Bernoulli trials with the same parameter $p$.

Since $\mathbb{V}(W) = \mathbb{E}(W)/(1-p)$, $\mathbb{E}(W) < \mathbb{V}(W)$, the negative binomial distribution is over-dispersed—as opposed to the under-dispersed binomial distribution, and the Poisson distribution which is neither over-dispersed nor under-dispersed.

In [8], the distribution of

$$S = \sum_{i=1}^{n} X_i, \tag{3}$$

with $X_i \sim \text{Bernoulli}(p_i)$, is approximated by $W \sim \text{NegBin}(r,p)$ such that $\mathbb{E}(W) = \mathbb{E}(S)$ i.e. $r$ and $p$ are selected so that

$$\frac{rp}{1-p} = \sum_{i=1}^{n} p_i.$$

Assuming the above, the total variation distance between $S$ and $W$ can be bounded from above by [8]:

$$\|S - W\| \leq \frac{(2-p)(1-p^r)}{r(1-p)} \sum_{j=1}^{n} \mathbb{E}(X_j)\, \mathbb{E}|W + U - W_j^*|, \tag{4}$$

where $U \sim \text{Geometric}(p)$ and, for each $j$, $W_j^*$ is a random variable with the same distribution as $W - X_j$ conditioned on having $X_j = 1$. Unfortunately, it is unclear how to interpret the later to compute or estimate the expected value of $|W + U - W_j^*|$ because $W$ imposes no restriction on $X_j$ but $W_j^*$ imposes that $X_j = 1$. In other words, the joint p.m.f. of $W$ and $W_j^*$ is unclear.

## 2 Review of the Chen-Stein Method for the Poisson Distribution

The Chen-Stein method is traditionally used for approximating the distribution of a sum of possibly dependent Bernoulli random variables by a Poisson random variable with the same mean as the sum.

The Chen-Stein method provides bounds for the total variation distance between the distributions of $S$ as in equation (3) and $Z \sim \text{Poisson}\big(\mathbb{E}(S)\big)$. It builds upon the following two results, whose proofs may be found in [4].

**Lemma 2.1.** *$Z \sim$ Poisson$(\lambda)$ if and only if $\mathbb{E}(\lambda g(Z+1) - Zg(Z)) = 0$, for all bounded function $g : \mathbb{N} \to \mathbb{R}$.*

**Lemma 2.2.** *If $f : \mathbb{N} \to \mathbb{R}$ is a bounded function such that $\mathbb{E}(f(Z)) = 0$ when $Z \sim$ Poisson$(\lambda)$, then there exists a bounded $g : \mathbb{N} \to \mathbb{R}$ such that:*

$$f(k) = \lambda g(k+1) - k g(k), \quad k \geq 0.$$

*The function $g$ is unique if one imposes that $g(0) = 0$.*

To see how the Chen-Stein method works, consider $A \subset \mathbb{N}$ and the function $f : \mathbb{N} \to \mathbb{R}$ defined as:

$$f(n) := [[n \in A]] - \mathbb{P}(Z \in A), \quad n \in \mathbb{N};$$

where $[[\cdot]]$ denotes the indicator function of the proposition within. Observe that:

$$\mathbb{E}\big(f(S)\big) = \mathbb{P}(S \in A) - \mathbb{P}(Z \in A).$$

On the other hand, since $\mathbb{E}(f(Z)) = 0$, Lemma 2.2 asserts the existence of a function $g_A$ such that:

$$\mathbb{E}(f(S)) = \mathbb{E}\big(\lambda\, g_A(S+1) - S\, g_A(S)\big).$$

Hence:

$$\mathbb{P}(S \in A) - \mathbb{P}(Z \in A) = \mathbb{E}\big(\lambda g_A(S+1) - S g_A(S)\big),$$

which implies that

$$\|S - Z\| \leq \sup_{A \subset \mathbb{N}} |\mathbb{E}\big(\lambda\, g_A(S+1) - S\, g_A(S)\big)|. \tag{5}$$

Observe that if $S$ had a Poisson distribution with parameter $\lambda$ then the right hand-side above would be zero. The key of the Chen-Stein method to approximate the distribution of $S$ by $Z$ is that if $S$ has an approximate Poisson distribution with rate $\lambda$ then the right hand-side above should be approximately zero.

## 3 Chen-Stein Method for the Negative Binomial

In this section, we state and show some results that are analogous to the ones used in applying the Chen-Stein method to the Poisson distribution. These are likely important steps for deriving an upper-bound for the total variation distance between a sum of Bernoulli's and a corresponding negative binomial.

In what follows, $W \sim \text{NegBin}(r,p)$. First we prove the following claim.

**Lemma 3.1.** *If $W \sim$ NegBin$(r,p)$ then $\mathbb{P}(W = 0) = (1-p)^r$, and*

$$p\,(w+r)\,\mathbb{P}(w) - (w+1)\,\mathbb{P}(w+1) = 0, \quad w \geq 0.$$

*Proof.* By definition

$$\mathbb{P}(W = 0) = \binom{r-1}{0}(1-p)^r\, p^0 = (1-p)^r.$$

On the other hand, for $w \geq 0$:

$$\begin{aligned}(w+1)\,\mathbb{P}(W = w+1) &= (w+1)\binom{w+r}{w+1}(1-p)^r\, p^{w+1} \\ &= p\frac{(w+r)!}{w!(r-1)!}(1-p)^r\, p^w \\ &= p(w+r)\frac{(w+r-1)!}{w!(r-1)!}(1-p)^r\, p^w \\ &= p\,(w+r)\binom{w+r-1}{w}(1-p)^r\, p^w \\ &= p\,(w+r)\,\mathbb{P}(w).\end{aligned}$$

□

Note that if for each $w \in \mathbb{N}$ we define the function $g_w(x) := [[x = w]]$, for all $x \in \mathbb{N}$, the previous result may be restated as saying that $\mathbb{E}\big(p\,(W+r)\,g_{w+1}(W+1) - W\,g_{w+1}(W)\big) = 0$. Therefore, the following result may be regarded as a more general version of Lemma 3.1.

**Lemma 3.2.** *If $W \sim$ NegBin$(r,p)$ and $g : \mathbb{N} \to \mathbb{R}$ is a bounded function, then*

$$\mathbb{E}\big(p\,(W+r)\,g(W+1) - W\,g(W)\big) = 0.$$

*Proof.* Observe that

$$\begin{aligned}
\mathbb{E}\big(p\,(W+r)\,g(W+1)\big) &= \sum_{w=0}^{\infty} p\,(w+r)\,g(w+1)\binom{w+r-1}{w}(1-p)^r\,p^w \\
&= \sum_{i=1}^{\infty} p\,(i+r-1)\,g(i)\binom{i+r-2}{i-1}(1-p)^r\,p^{i-1} \\
&= \sum_{i=1}^{\infty} g(i)\,\frac{(i+r-1)!}{(i-1)!(r-1)!}\,(1-p)^r\,p^i \\
&= \sum_{i=0}^{\infty} i\,g(i)\binom{i+r-1}{i}(1-p)^r\,p^i \\
&= E\big(W\,g(W)\big),
\end{aligned}$$

where for the second identity we have substituted $w = i - 1$. □

We can now prove the following result, which is analogous to Lemma 2.2.

**Corollary 3.3.** *Let $W$ be random variable that can only take values in $\mathbb{N}$. Then, $W \sim$ NegBin$(r,p)$ if and only if $\mathbb{E}\big(p\,(W+r)\,g(W+1) - W\,g(W)\big) = 0$, for all bounded function $g : \mathbb{N} \to \mathbb{R}$.*

*Proof.* Due to Lemma 3.2, it suffices to show that if $\mathbb{E}\big(p\,(W+r)\,g(W+1) - W\,g(W)\big) = 0$ for all $g : \mathbb{N} \to \mathbb{R}$ bounded, then $W \sim \text{NegBin}(r,p)$. For this fix a $w \in \mathbb{N}$ and define $g(x) := [[x = w+1]]$, for $x \in \mathbb{N}$. Since $g$ is bounded, we have:

$$\begin{aligned}
0 &= \mathbb{E}\big(p\,(W+r)\,g(W+1) - W\,g(W)\big) \\
&= \mathbb{E}\big(p\,(W+r)[[W = w]]\big) - \mathbb{E}\big(W\,[[W = w+1]]\big) \\
&= p\,(w+r)\,\mathbb{P}(W = w) - (w+1)\,\mathbb{P}(W = w+1),
\end{aligned}$$

i.e.

$$\mathbb{P}(W = w+1) = \frac{p\,(w+r)}{w+1}\mathbb{P}(W = w), \text{ for all } w \in \mathbb{N}.$$

Repeated applications of the above recursion for the p.m.f. of $W$ then gives that

$$\mathbb{P}(W = w) = \binom{w+r-1}{w}p^w\,\mathbb{P}(W = 0), \text{ for all } w \in \mathbb{N}.$$

Finally, since we must have

$$\sum_{w=0}^{\infty} \mathbb{P}(W = w) = 1,$$

the above identity implies that

$$\mathbb{P}(W = 0) = \frac{1}{\sum\limits_{w=0}^{\infty} \binom{w+r-1}{w} p^w} = \frac{(1-p)^r}{\sum\limits_{w=0}^{\infty} \binom{w+r-1}{w} p^w (1-p)^r} = (1-p)^r,$$

where for the last identity we have used that $\binom{w+r-1}{w} p^w (1-p)^r$ is the p.m.f. of a negative binomial distribution. As a result:

$$\mathbb{P}(W = w) = \binom{w+r-1}{w} p^w (1-p)^r, \text{ for all } w \in \mathbb{N};$$

which shows that $W$ has a negative binomial distribution, as claimed. □

Finally, we prove a result that is analogous to Lemma 2.2 but for the negative binomial distribution.

**Corollary 3.4.** *If $f : \mathbb{N} \to \mathbb{R}$ is a bounded function such that $\mathbb{E}(f(W)) = 0$, when $W \sim NegBin(r,p)$, then there exists a bounded function $g : \mathbb{N} \to \mathbb{R}$ such that:*

$$f(w) = (1-p)\,(w+r)\,g(w+1) - w\,g(w), \text{ for all } w \in \mathbb{N}. \tag{6}$$

*Furthermore, $g$ is unique if one imposes that $g(0) = 0$.*

*Proof.* To determine the function $g$, assume that the identity in equation (6) is satisfied. Multiplying both sides of this equation by $(1-p)^{w-1}(w+r-1)!/w!$, we obtain for $n \geq 2$ that

$$\begin{aligned}\sum_{w=1}^{n-1} (1-p)^{w-1} \frac{(w+r-1)!}{w!} f(w) &= \sum_{w=1}^{n-1} \left\{ (1-p)^w \frac{(w+r)!}{w!} g(w+1) - (1-p)^{w-1} \frac{(w-1+r)!}{(w-1)!} g(w) \right\} \\ &= (1-p)^{n-1} \frac{(n+r-1)!}{(n-1)!} g(n) - r!\,g(1) \\ &= (1-p)^{n-1} \frac{(n+r-1)!}{(n-1)!} g(n) - \frac{(r-1)!}{1-p} f(0),\end{aligned}$$

where the second to last step is justified by telescopic summation, and the last step is due to the fact that $f(0) = (1-p)\,r\,g(1)$. Thus:

$$g(n) = \frac{(n-1)!}{(n+r-1)!\,(1-p)^{n-1}} \sum_{w=0}^{n-1} (1-p)^{w-1} \frac{(w+r-1)!}{w!} f(w), \text{ for } n \geq 1.$$

Clearly, the steps that led to this identity are reversible; in particular, the function $g$ not only solves equation (6) but it is also uniquely defined for $n \geq 1$. Furthermore, if we interpret $\sum\limits_{w=0}^{-1} \equiv 0$, the above is the only function $g$ such that $g(0) = 0$.

It remains to show that $g$ is bounded. For this, recall that $\mathbb{E}\big(f(W)\big) = 0$. In particular:

$$\sum_{w=0}^{\infty} (1-p)^{w-1} \frac{(w+r-1)!}{w!} f(w) = 0,$$

which implies that

$$g(n) = -\frac{(n-1)!}{(n+r-1)!\,(1-p)^{n-1}} \sum_{w=n}^{\infty} (1-p)^{w-1} \frac{(w+r-1)!}{w!} f(w)$$
$$= -\frac{(n-1)!}{(n+r-1)!} \sum_{k=0}^{\infty} (1-p)^{k} \frac{(k+n+r-1)!}{(k+n)!} f(k+n).$$

As a result:

$$|g(n)| \le \sup_{w\in\mathbb{N}} |f(w)| \cdot \frac{(n-1)!}{(n+r-1)!} \cdot \sum_{k=0}^{\infty} (1-p)^{k} \frac{(k+n+r-1)!}{(k+n)!}. \tag{7}$$

But note the following about the ratio of consecutive terms in the above sum:

$$\frac{(1-p)^{k+1} \frac{(k+1+n+r-1)!}{(k+1+n)!}}{(1-p)^{k} \frac{(k+n+r-1)!}{(k+n)!}} = (1-p)\frac{k+n+r}{k+n+1} \le (1-p)\frac{n+r}{n+1},$$

where we have used that the transformation $k \to (k+n+r)/(k+n+1)$ is a decreasing function of $k \ge 0$. Since $\lim_{n\to\infty} \frac{n+r}{n+1} = 1$ and $(1-p) \le (1-p/2)$, the argument of the well-know ratio-test for infinite series implies that for all $n$ large enough:

$$|g(n)| \le \sup_{w\in\mathbb{N}} |f(w)| \cdot \frac{1}{n} \cdot \sum_{k=0}^{\infty} \left(1 - \frac{p}{2}\right)^{k}.$$

Since the right hand-side above tends to 0 as $n$ tends to infinity, the function $g$ is bounded, which completes the proof of the lemma. □

This project did not determine if it is possible to combine the lemmas in this section to obtain upper-bounds similar to the ones in [4] for the Poisson approximation of a sum of Bernoulli random variables.

# 4 Performance on Numerical Experiments

In what follows

$$S = \sum_{i=1}^{n} X_i,$$

where $X_1, \ldots, X_n$ are Bernoulli random variables, while

$$W \sim \text{NegBin}(r, t)$$

is a negative binomial random variable. In this section we examine various numerical experiments that range in terms of the dependency between the Bernoulli random variables, and examine how the distribution of $W$ approximates that of $S$ for suitable values of the parameters $r$ and $t$ such that $\mathbb{E}(S) = \mathbb{E}(W)$.

## 4.1 Independent and Identically Distributed Bernoullis

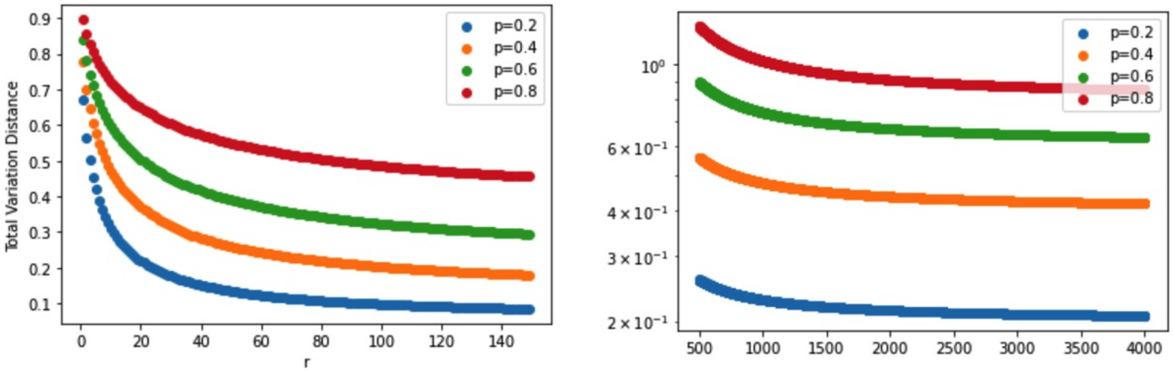
Let $q = 1-p$. Let $n = 100$ and consider the case where $X_1, \ldots, X_n$ are i.i.d. Bernoulli$(p)$ random variables. In particular, $S \sim$ Binomial$(100, p)$. From [8]:

$$\|S - W\| \le \frac{(2-q)(1-q^r)}{r(1-q)} np \left(p + \frac{1-q}{q}\right).$$

In Figure 1, we observe how the TVD changes as a function of $r$ when $p \in \{0.2, 0.4, 0.6, 0.8\}$. As can be seen, in all cases the total variation distance or its theoretical upper-bound appear both to decrease to an asymptotic value.

The theoretical upper-bound for the total variation distance, which is depicted in Figure 1(b), is above 1 for $r < 500$. However, as $r$ increases, the bound seems to decrease slowly towards the corresponding value of $p$.

In Figure 2(a)-(d), we examine and compare the exact p.m.f.'s that correspond to the cases in Figure 1(a). As can be seen, in all cases, the approximation in Figure 2(a) is much better than the approximation in Figure 2(b), which is in turn better than the approximation in Figure 2(c), and so on.



(a)

(b)

Figure 1: **Comparison of the TVD for identically and independently distributed Bernoullis.** Total variation distance between $S$ and $W$ as a function of $r$ when $n = 100$, and $p = 0.2$ (blue), $0.4$ (orange), 0.6 (green), and 0.8 (red). (a): observed total variation distance between $S$ and $W$. (b): theoretical upper-bound for total variation distance between $S$ and $W$.

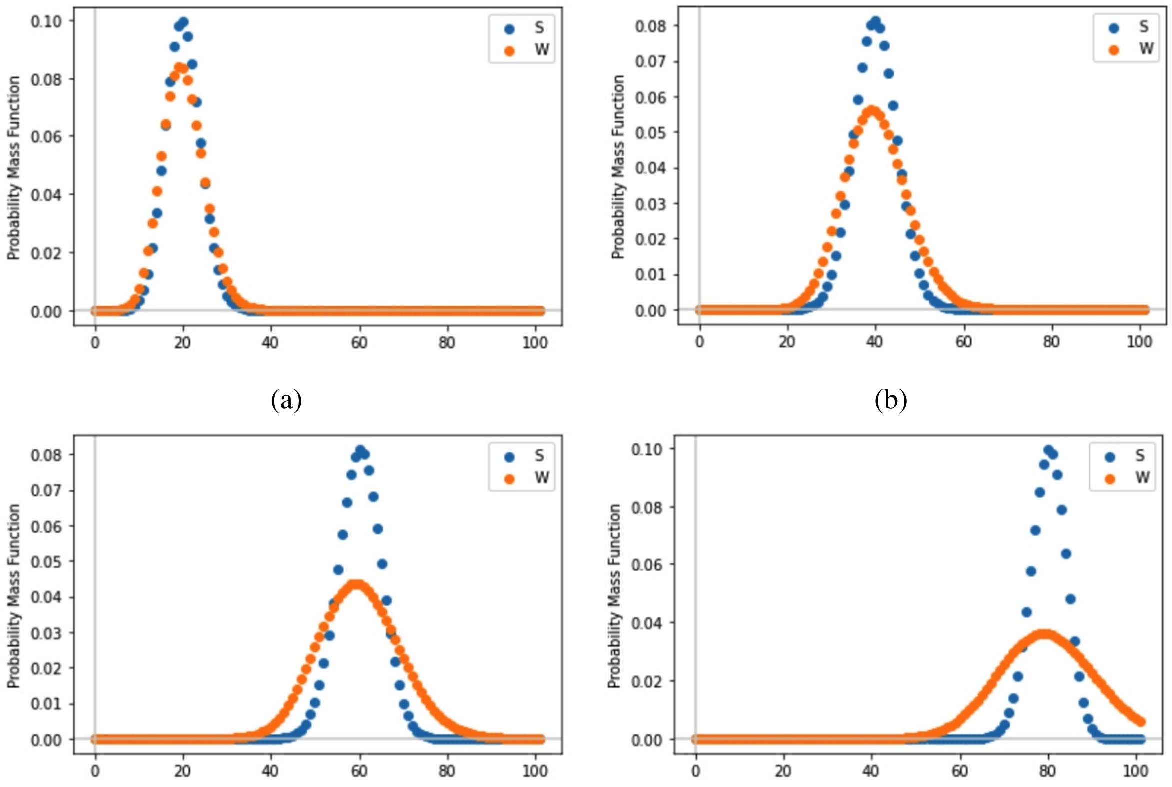
## 4.2 Independent but Not Identically Distributed Bernoullis

Again consider the case with $n = 100$ but now let $X_1, ...X_n$ be independent Bernoulli random variables such that $X_i \sim \text{Bernoulli}(p_i)$ with $p_i \in [0, 1]$. These parameters were selected at random. We display their histogram in Figure 3(c).

Let $q = 1 - p$. Using the formula from [8], we end up with the following upper bound:

$$\|S - W\| \leq \frac{(2-q)(1-q^r)}{r(1-q)} \sum_{j=1}^{n} p_j \left( p_j + \frac{1-q}{q} \right).$$

As can be seen in Figure 3(a), the TVD appears to continuously decrease as $r$ increases, seemingly converging to an asymptotic value of approximately $0.3$, which is less accurate than the minimum total variation distance encountered in Section 4.1 and Section 4.4, but more accurate than the minimum distance encountered in Section 4.3. Figure 3(d) compares the p.m.f.'s of $S$ and $W$ for when $r = 140$. It is clear from the plot that the p.m.f.'s appear rather different, despite retaining the same shape and center. As of the plot of the theoretical upper-bound for total variation distance between $S$ and $W$ as a function of $r$, it appears to converge towards an asymptotic value of roughly 0.65. Thus, overall it can be concluded that the theoretical upper-bound is a huge overestimate and unlikely to be useful in practice.



(a)

(b)

(c)

(d)

Figure 2: **Comparison of the distribution of S and W for identically and independently distributed Bernoullis.** The plots display the p.m.f. of $S$ (blue) and $W$ (orange) when $n = 100, r = 150$, and (a) $p = 0.2$, (b) $p = 0.4$, (c) $p = 0.6$ and (d) $p = 0.8$.

### 4.3 Identically Distributed Bernoullis with Constant Correlation

Let $c \neq 0$ be a real constant. In this section, we consider the case where $X_1, .., X_n$ are identically distributed Bernoulli random variables but $\text{cov}(X_i, X_j)$ is non-zero and constant for all $i \neq j$.

To show that the above model is feasible, consider $X \sim \text{Bernoulli}(p)$. If $X = 1$, define $Z_1 = \ldots = Z_n = 0$. Instead, if $X = 0$, let $Z_1, \ldots, Z_n$ be i.i.d Bernoulli$(q)$. Define

$$X_i := X + Z_i, \text{ for } i = 1, \ldots, n.$$

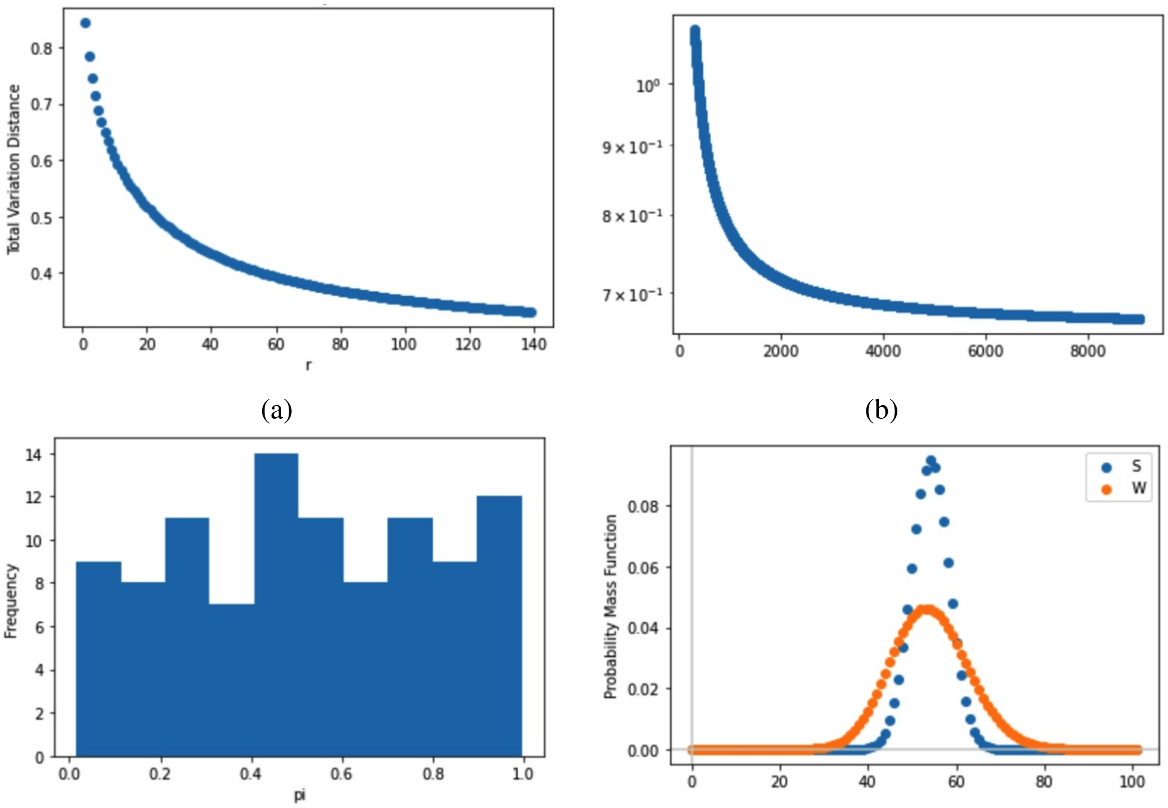
We first show that $X_1, ..., X_n$ are equally distributed, noting from the construction that

$$\begin{aligned} \mathbb{P}(X_i = 0) &= \mathbb{P}(Z_i = 0|X = 0) \cdot \mathbb{P}(X = 0) = (1 - q)(1 - p); \\ \mathbb{P}(X_i = 1) &= 1 - \mathbb{P}(X_i = 0) = p + q - pq. \end{aligned}$$

So $X_1, \ldots, X_n$ are equally distributed Bernoulli random variables with parameter $p + q - pq$.

On the other hand:

$$\begin{aligned} \mathbb{E}(X_i X_j) &= \mathbb{P}(X_i = X_j = 1) \\ &= \mathbb{P}(Z_i = Z_j = 1|X = 0) \cdot \mathbb{P}(X = 0) + \mathbb{P}(Z_i = Z_j = 0|X = 1) \cdot \mathbb{P}(X = 1) \\ &= q^2(1 - p) + p. \end{aligned}$$



(a)

(b)

(c)

(d)

Figure 3: **Comparison of the distribution of S and W in the independent but not identically distributed case.** (a): experimental total variation distance between $S$ and $W$ as a function of $r$ when $n = 100$ and the parameter $p_i \in [0,1]$ of all the Bernoullis were randomized. (b): theoretical total variation distance between $S$ and $W$ as a function of $r$ when $n = 100$. (c): distribution of randomized $p_i$'s. (d): p.m.f. of $S$ (blue) and $W$ (orange) when $n = 100$, and $r = 53$.

As a result, for $i \neq j$:

$$\operatorname{cov}(X_i, X_j) = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i) \cdot \mathbb{E}(X_j) = q^2(1-p) + p - \big(p + q - pq\big)^2,$$

which does not depend on $i$ and $j$, as claimed.

Recall that $S = \sum_{i=1}^{n} X_i$. The p.m.f. of $S$ is:

$$\mathbb{P}(S = s) = \begin{cases} (1-p)\binom{n}{s} q^s (1-p)^{n-s} & , s = 0, \ldots, n-1; \\ (1-p)\, q^n + p & , s = n. \end{cases}$$

We can use the above to compute explicitly the total variation distance between $S$ and $W \sim \text{NegBin}(r, t)$ where $(r, t)$ are such that $\mathbb{E}(S) = \mathbb{E}(W)$.

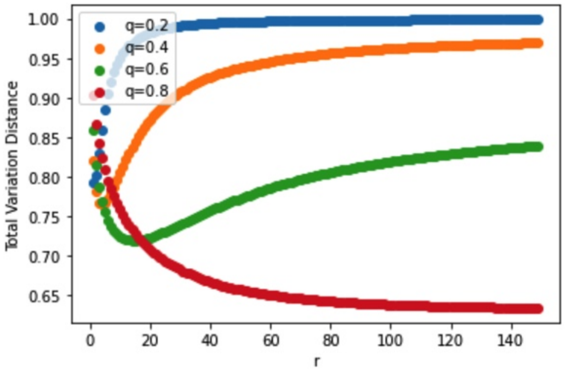


Figure 4: **Comparison of the total variation distance for summation of identically distributed Bernoullis with a constant correlation**. Total variation distance between $S$ and $W$ as a function of $r$ when $n = 100$, $p = 0.5$, and $q \in \{0.2, 0.4, 0.6, 0.8\}$.

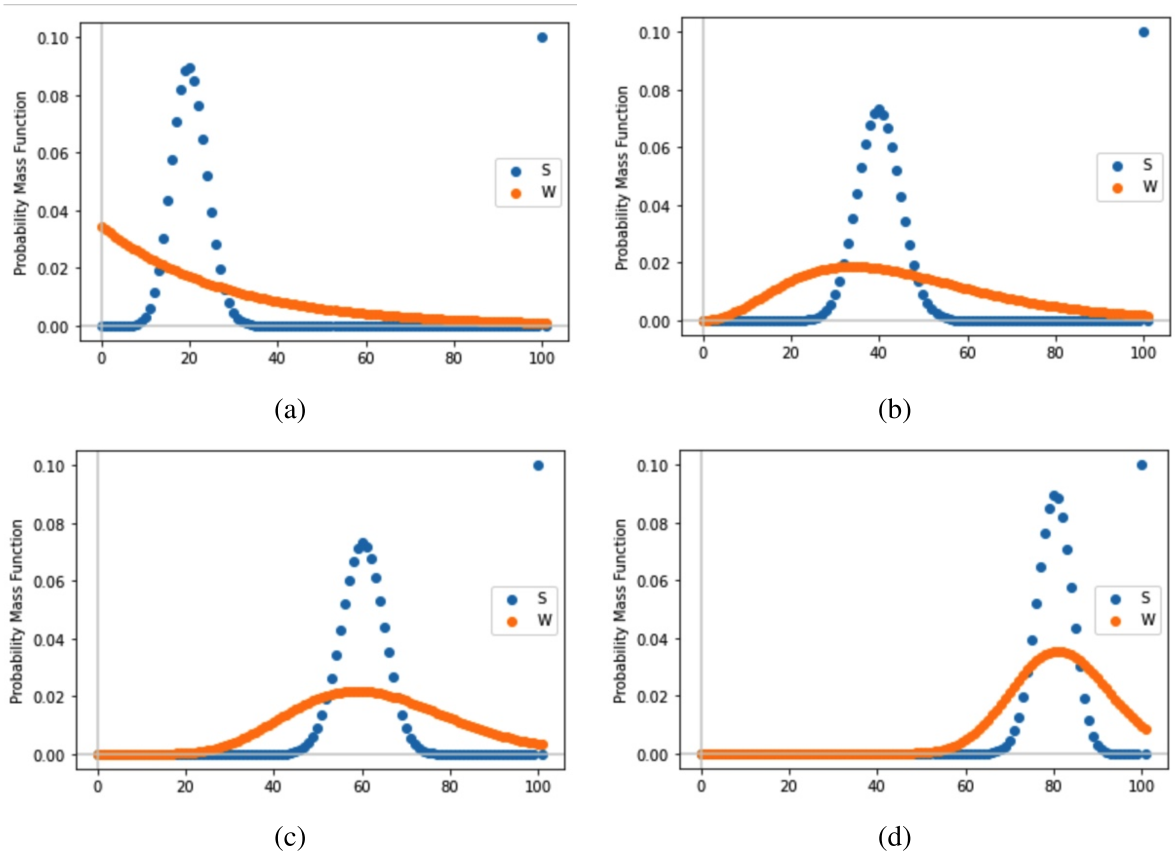
In Figure 4, we observe how the total variation distance changes as a function of $r$ when $p = 0.5$ and $q \in \{0.2, 0.4, 0.6, 0.8\}$. As can be seen, when $q \in \{0.2, 0.4, 0.6\}$ the total variation distance has a local minimum, and eventually increases and seemingly reaches an asymptotic value. For $q = 0.2, 0.4, 0.6$ the minimum total variation distance can be observed at $r = 1, 4, 14$ and takes on the values 0.792, 0.765, and 0.719, respectively. However, when $q = 0.8$, the total variation distance continues to decrease, seemingly approaching an asymptotic value of about 0.63. Through various other experiments, we have observed that regardless of how $q$ and $p$ vary, all plots display one of the two types of behavior just described. We note that even as $q$ approaches 1, the total variation distance seems to approach the asymptotic value of 0.6 approximately—though never lower. The reason for these large total variation distance values, and as seen in Figure 5(a)-(d), is that even for the value of $r$ that that minimizes $\|S - W\|$ the p.m.f. of $W$ is a bad approximation of the p.m.f. of $S$.

## 4.4 Bernoullis with a Markovian Dependence

Our final numerical example considers a stationary Markov chain $(X_n)_{i=1}^{n}$ with state space $\{0, 1\}$. The motivation for this is that, because the state space is finite, the dependence between $X_i$ and $X_j$ decays exponentially fast with $|i - j| > 0$. This contrasts with the previous experiment where the correlation remained fixed.

The probability transition matrix (p.t.m.) of the chain is given by

$$M := \begin{pmatrix} 1-\alpha & \beta \\ \alpha & 1-\beta \end{pmatrix},$$



(a)

(b)

(c)

(d)

Figure 5: **Comparison of the distribution of S and W for identically distributed Bernoullis with a constant correlation.** The plots display the p.m.f. of $S$ (blue) and $W$ (orange) when $n = 100$, $p = 0.5$ and (a) $q = 0.2$, $r = 1$, (b) $q = 0.4$, $r = 4$, (c) $q = 0.6$, $r = 14$ and (d) $q = 0.8$, $r = 150$.

which we have represented visually in Figure 6. It is assumed from now on that $(\alpha + \beta) > 0$.

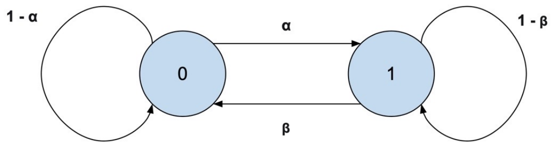


Figure 6: **Visual representation of Markov chain with state space $\{\mathbf{0}, \mathbf{1}\}$.**

A simple calculation shows that the stationary distribution of the chain is

$$\pi := \left( \frac{\beta}{\alpha + \beta}, \frac{\alpha}{\alpha + \beta} \right).$$

In particular, if $\mathbb{P}(X_0 = 0) = \pi(0)$ and $\mathbb{P}(X_0 = 1) = \pi(1)$ then $X_1, \ldots, X_n$ are identically distributed Bernoulli($\pi(1)$) random variables. However, these are dependent random variables though their dependence decays exponentially fast. To see this, note the diagonalization of the p.t.m.:

$$M = \begin{pmatrix} \beta/\alpha & -1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 - \alpha - \beta \end{pmatrix} \cdot \begin{pmatrix} \pi(1) & \pi(1) \\ -\pi(1) & \pi(0) \end{pmatrix}.$$

In particular:

$$M^n = \begin{pmatrix} \beta/\alpha & -1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & (1-\alpha-\beta)^n \end{pmatrix} \cdot \begin{pmatrix} \pi(1) & \pi(1) \\ -\pi(1) & \pi(0) \end{pmatrix},$$

i.e. for $0 \le i < j \le n$:

$$\begin{aligned} \mathbb{P}(X_j = 0 | X_i = 0) &= \pi(0) + \pi(1) \cdot (1-\alpha-\beta)^{j-i}; \\ \mathbb{P}(X_j = 1 | X_i = 0) &= \pi(1) - \pi(1) \cdot (1-\alpha-\beta)^{j-i}; \\ \mathbb{P}(X_j = 0 | X_i = 1) &= \pi(0) - \pi(0) \cdot (1-\alpha-\beta)^{j-i}; \\ \mathbb{P}(X_j = 1 | X_i = 1) &= \pi(1) + \pi(0) \cdot (1-\alpha-\beta)^{j-i}. \end{aligned}$$

Observe that in this case

$$S = \sum_{i=1}^{n} X_i$$

represents the number of visits to state ① in the first $n$ transitions of the chain. One should expect $S$ to be well-approximated by a negative binomial because this distribution can be expressed as a sum of independent geometric random variables and, each time state ① is visited from ⓪, the total number of visits to ① is geometric. Nevertheless, the distribution of $S$ is not negative binomial because the number of visits to ① from ⓪ is random, and because $S$ keeps tracks of the visits to state ① only up to time $n$.

In Figure 7, we observe how the total variation distance between $S$ and $W$ changes as a function of $r$ when $\beta = 0.5$ and $\alpha \in \{0.2, 0.4, 0.6, 0.8\}$. As seen in the figure, when $\alpha = 0.2$, the total variation distance seemingly reaches a minimum of around 0.0117 at $r = 89$. When $\alpha = 0.4, 0.6, 0.8$, the distance continues to decrease with $r$ and apparently reaches an asymptotic value of around 0.155, 0.3, 0.43, respectively. We had observed similar behaviours in Section 4.3, however, there is a key difference: while the approximation in Section 4.3 tended to improve as $q$ increased, in this case the approximation worsens as $\alpha$ increases. Nevertheless, the total variation distance appears to approach 0 as $\alpha$ and $\beta$ also approach 0, making this a much more suitable approximation.

In Figure 8(a)-(d), we compare the p.m.f. of $S$ with the one of $W$ associated with the parameter $r$ that minimizes $\|S - W\|$. As can be seen in the figure, the approximation worsens as $\alpha$ increases. Through various other experiments, it appears that the total variation distance is at a minimum when both $\alpha$ and $\beta$ approach 0, but as either $\alpha$ or $\beta$ increase, the distance increases dramatically.

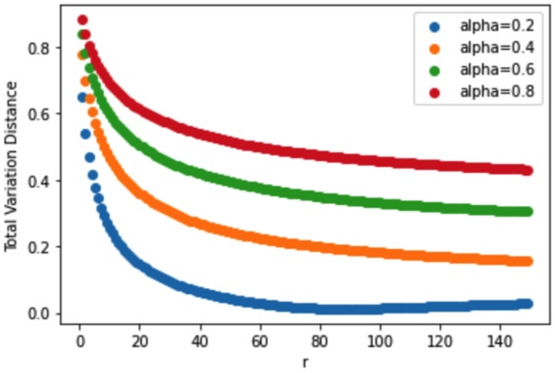
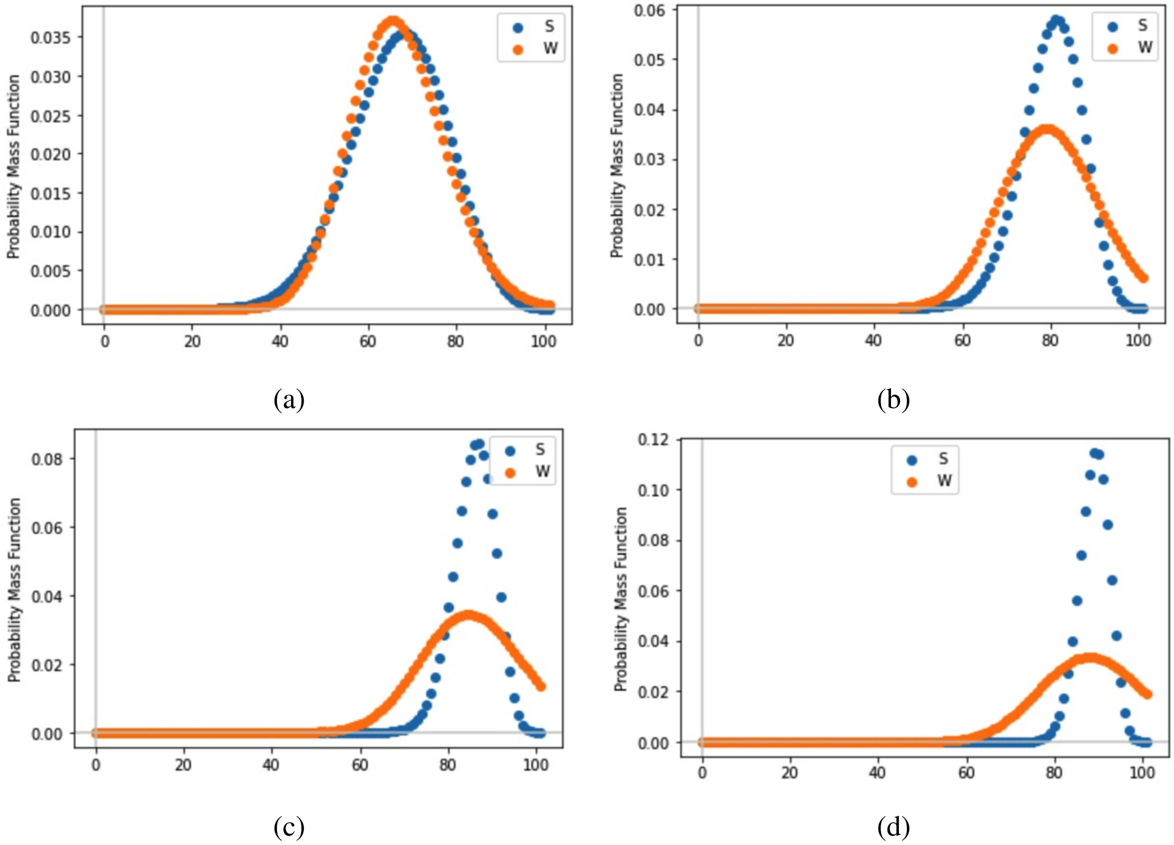


Figure 7: **Comparison of the TVD for Bernoullis in a Markov Chain**. Total variation distance between $S$ and $W$ as a function of $r$ when $\beta = 0.5, n = 100$, and $\alpha \in \{0.2, 0.4, 0.6, 0.8\}$.



(a)

(b)

(c)

(d)

Figure 8: **Comparison of the distribution of S and W for Bernoullis in a Markov chain.** The plots display the p.m.f. of $S$ (blue) and $W$ (orange) when $n = 100$, $\beta = 0.5$, and (a) $\alpha = 0.2$, $r = 89$, (b) $\alpha = 0.4$, $r = 150$, (c) $\alpha = 0.6$, $r = 150$, and (d) $\alpha = 0.8$, $r = 150$.

## 5 Conclusions and Future Work

In this project, we have stated and proved various lemmas related to finding a bound for the total variation distance between a sum of Bernoulli's and a corresponding negative binomial. We could not deduce an explicit and practical upper bound for the distance, though.

On the other hand, the upper-bound proposed in [8]—see equation (4), has an ambiguous interpretation that renders it impractical in some situations. Thus, we proceeded to explore the exact total variation distance in various experimental cases, which vary in how correlated the Bernoulli's in the summation are.

When computing the total variation distance between $S$ and $W$ as a function of $r$, we observed two types of behaviours. In one type, the distance starts out by decreasing to minimum value before increasing to a seemingly asymptotic one. In the other type of behavior, the distance seems to continuously decrease to certain asymptotic value. Overall, it appears that the theoretical upper-bound of the total variation distance is overly conservative. In fact, it can sometimes be much greater than one. Furthermore, the total variation distance may be quite high; rendering the approximation of a sum of Bernoulli's by a negative binomial often impractical.

For future work, it would be significant to correct or clarify the bound of total variation distance found in [8], and possibly deriving a different bound. In addition, it would be useful to apply the approximation of over-dispersed sums of Bernoulli random variables to real-world situations.

## References

[1] Louis H. Y. Chen. Poisson approximation for dependent trials, June, 1975. Ann. Probab. 3 (3) 534-545.

[2] Marek et al. Gierliński. Statistical models for rna-seq data derived from a two-condition 48-replicate experiment. *Bioinformatics*, 31(22):3625–30, 2015.

[3] E.S. Lander and M.S. Waterman. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics*, 2:231–239, 1988.

[4] Manuel Lladser. Random graphs lecture notes, 2021. APPM 5565, University of Colorado Boulder.

[5] Pieta Schofield et al Marek Gierlinski, Christian Cole. Statistical models for RNA-seq data derived from a two-condition 48-replicate experiment, 2015. Bioinformatics, 31(22):3625–3630.

[6] Louis Gordon Richard Arratia, Larry Goldstein. Poisson approximation and the Chen-Stein method, 1990. Statistical Science, Vol. 5, No. 4, 403-434.

[7] C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables., 1972. Proceedings of the Sixth Berkeley Symposium on mathematical statistics and probability, 583-602.

[8] M. J. Phillips Timothy C. Brown. Negative binomial approximation with Stein's method, 1999. Methodology and Computing in Applied Probability, 1:4, 407-421.