Zero Bias Enchanced Stein Couplings

Larry Goldstein*

Abstract

The Stein couplings of Chen and Roellin [6] vastly expanded the range of applications for which coupling constructions in Stein’s method for normal approximation could be applied, and subsumed both Stein’s classical exchangeable pair, as well as the size bias coupling. A further simple generalization includes zero bias couplings, and also allows for situations where the coupling is not exact. The zero bias versions result in bounds for which often tedious computations of a variance of a conditional expectation is not required. An example to the Lightbulb process shows that even though the method may be simple to apply, it may yield improvements over previous results that had achieved bounds with optimal rates and small, explicit constants.

Keywords: Coupling; Normal Approximation, Zero Bias.

MSC2020 subject classifications: 60F05; 62P10; 60C05.

1 Introduction

Stein’s method was introduced in the seminal paper [29], and has since proved to be an extraordinarily useful tool in the area of normal approximation, see the text of [5] and references therein, and an impressive, constantly expanding array of other distributions, [4], [22], [25], [3], [26], [15], [11], as well as in concentration inequalities [1], [2], [12], [18] stochastic analysis [24], [23], [19], and data science [9], [20], [21], [10].

The method proceeds by the use of a characterizing equation for some target distribution which is to serve as an approximation to that of a random variable $W$ of interest. In the case of the normal, $W$ has the $N(\mu, \sigma^2)$ distribution if and only if

$$E[(W - \mu) f(W)] = \sigma^2 E[f'(W)]$$

for all $f \in F$, (1.1)

where here, and in like displays that follow in this section, we implicitly take $F$ to be the class of functions for which the quantities written exist, which in particular will always include the collection of infinitely differentiable functions with compact support. Next, given any collection of functions $\mathcal{H}$, we may define the pseudo-metric

$$d_\mathcal{H}(X, Y) = \sup_{h \in \mathcal{H}} |Eh(Y) - Eh(X)|,$$

(1.2)

which, for example, produces the Wasserstein $d(\cdot, \cdot)$ and Kolmogorov metric $d_K(\cdot, \cdot)$ by taking $\mathcal{H}$ to be the class

$$\text{Lip}_1 = \{h : |h(y) - h(x)| \leq |y - x|, \{x, y\} \subset \mathbb{R}\} \text{ or Ind} = \{h : h(x) = 1_{(-\infty, z]}(x), z \in \mathbb{R}\},$$

*Department of Mathematics, University of Southern California, Los Angeles, CA USA E-mail: larry@math usc edu
Enhanced Stein Couplings

of Lipschitz-1 functions, or the collection of all indicators of right closed, left infinite intervals, respectively.

Now consider the normal approximation of a random variable $W$ with mean $\mu$ and finite, non-zero variance $\sigma^2$, which without loss of generality we may take to be 1. Given $h$ from the class of test functions $H$, and in view of the characterizing equation (1.1), with $Z \sim N(0, 1)$ and $Nh = E[h(Z)]$, we let $f_h$ be the unique bounded solution to the Stein equation

$$f'(w) - (w - \mu)f(w) = h(w) \quad \text{thus yielding}$$

$$E[f'_h(W) - (W - \mu)f_h(W)] = E[h(W) - Nh]. \quad (1.3)$$

One then obtain bounds for the quantity on the right hand side of (1.2) in terms of the left hand side of (1.3), to wit,

$$d_H(W, Z) = \sup_{h \in H} |E[f'_h(W) - (W - \mu)f_h(W)]|, \quad (1.4)$$

which, to be put in a usable form, will ultimately involve bounds involving the solution $f_h$, such as on the magnitude of its derivatives. Though the form (1.4) may seem at first glance to be more complicated and more difficult to handle than (1.2), the former involves the distribution of only a single random variable.

It is perhaps for this reason that one of the best known ways to handle the left hand side of (1.3) is through coupling constructions. An advance in understanding such coupling methods was achieved in [6] by the introduction of the family of Stein couplings, of which many previous couplings, such as the exchangeable pair, and size bias couplings, are special cases.

We say the triple of random variables $(W', W, G)$ is a Stein coupling when

$$E[(W - \mu)f(W)] = E[G(f(W') - f(W))]. \quad (1.5)$$

When the mean of $W$ exists the case $f$ identically equal to 1 shows $E[W] = \mu$. To illustrate one such coupling, with $W$ having mean zero, when an exchangeable pair $(W', W)$ satisfies the linear regression condition

$$E[W' - W|W] = (1 - \lambda)W \quad (1.6)$$

for some $\lambda \in (0, 1)$, we obtain Stein’s classical exchangeable pair that satisfies

$$E[Wf(W)] = \frac{1}{2\lambda}E[(W' - W)(f(W') - f(W))], \quad (1.7)$$

that is, $(W', W, (W' - W)/2\lambda)$ is a Stein coupling.

Further, recalling that for a random variable $W \geq 0$ with positive mean $\mu$, we say $W^*$ has the $W$-size bias distribution when

$$E[Wf(W)] = \mu E[f(W^*)]. \quad (1.8)$$

The size bias distribution gives rise to a Stein coupling by taking the triple to be $(W^*, W, \mu)$, with the sized biased variable $W^*$ defined on the same space as $W$. A coupling of a rather different type from the two just described is Coupling 2A given in [6], which demonstrates the large range of possibilities covered.

One key step in computing bounds in the Kolmogorov or Wasserstein metric to the normal using Stein couplings requires taking the expected absolute value, or a variance of a conditional expectation, the latter typically of a form such as

$$\text{Var}E \left((W' - W)^2|W\right) \quad \text{or} \quad \text{Var}E (W' - W|W); \quad (1.9)$$
see, for instance, the terms $r_1, r_2$ and $r_3$ in Theorem 2.1 of [6] for the Kolmogorov bound, or Theorem 3.20 of [28], for the Wasserstein. However, these terms are often difficult, tedious or even not tractable to compute. For instance, computing bounds using size bias for the lightbulb problem, described in detail below, the computation of the second term in [17] required a delicate and highly detailed analysis of the eigenvalues of a certain Markov chain of multinomial type.

On the other hand, the zero bias coupling does not require the computation of variances as in (1.9), but rather expectations that can be more easily evaluated. We recall that from [16] (see also [5]) that for a mean $\mu$ random variable $W$ with variance $\sigma^2 \in (0, \infty)$, we say that $W^*$ has the $W$-zero bias distribution when
\begin{equation}
E[(W - \mu)f(W)] = \sigma^2 E[f'(W^*)],
\end{equation}
where we have subtracted the mean as in [7]. Taking $\mu = 0$ and $\sigma^2 = 1$ for simplicity, we obtain the bound
\begin{equation}
d(Z, W) \leq 2d(W, W^*)
\end{equation}
in Wasserstein distance $d$ between $W$ and the standard normal $Z$, see [13] and [14], the right hand which can be upper bounded by $2E|W^* - W|$ for any coupling of $(W^*, W)$, in contrast to (1.9). Kolmogorov bounds to the normal can be provided by a bound on $|W^* - W|$, see [5]. Theorem 2.3 below provides bounds that generalize results previously obtained in both these metrics in the framework of ‘zero bias enhanced Stein couplings’.

Zero biased enhanced Stein couplings, or zbest for short, include both Stein and zero bias couplings as special cases. For a random variable $W$ with mean $\mu$ we say the triple $(W'', W', G)$ of random variables is a zbest coupling when for all bounded continuous functions $f$
\begin{equation}
E[(W - \mu)f(W)] = E[G(f(W'')) - f(W')].
\end{equation}
In contrast to (1.5) neither $W''$ nor $W'$ need to be equal to, nor even have the same distribution, as the random variable $W$. Relaxing that constraint in (1.5) leads to some additional, useful flexibility. Though the notational use of the same symbol for expectation on both sides of (1.11) implies that the triple $(W'', W', G)$ and $W$ are constructed on the same space, in principle one can introduce a triple and compare the right hand of (1.11) to expectations on the left that depend only on $f$ and the distribution of $W$.

Clearly the class of zbest couplings include Stein couplings, and we show in the next section that the generalization (1.11) produces a family of couplings that include the zero bias coupling as a special case. In addition, in Lemma 2.1 below, we show how starting with a Stein coupling one may be naturally led to a ‘second order’ zbest coupling for which quantities such as (1.9) may not arise when computing bounds to the normal. Naturally, nothing being for free, the work that would otherwise be required for the computation of (1.9) must be paid for in another manner. This work involves the construction of the secondary coupling that arises from an initial Stein coupling. The advantage of this approach is that the further coupling may be effected more easily than the computations needed for (1.9), and moreover, that this route may lead to better bounds.

We apply the methods developed here for the lightbulb problem first analyzed in [27] and further in [17]. In this instance demonstrates the computation of Kolmogorov and Wasserstein bounds is made easy, and leads to bounds that are superior to those produced by methods that require computation of quantities such as those in (1.9).

We present zbest couplings and the resulting bounds for normal approximation in the Wasserstein and Kolmogorov metric in Section 2 and the application to the Lightbulb Problem in Section 3.
Chapter 2: Zbest couplings

Throughout let $W$ be a mean $\mu$ random variable with finite, non-zero variance $\sigma^2$. Let $(W'', W', G)$ be a zbest coupling, that is, a triple satisfying (1.11). Set

$$D = W'' - W' \quad \text{and} \quad W^* = UW'' + (1 - U)W',$$

where $U$ is a standard uniform variable, independent of $W'', W'$ and $G$. For a smooth function $f$ for which the following expectations exist, and arbitrarily defining any $0/0$ expressions as $1$, integrating over the uniform variable we obtain

$$G(f(W'')) - f(W') = GD \left[ \frac{f(W'') - f(W')}{W'' - W'} \right] = E[DGf'(W^*)|W'', W', G].$$

Taking expectation and recalling (1.11) results in the identity

$$E[(W - \mu)f(W)] = E[G(f(W'')) - f(W')] = E[DGf'(W^*)].$$

Specializing to $f(w) = w$ in (2.2) we obtain

$$E[DG] = \sigma^2,$$

showing in particular that the product $DG$ cannot be zero almost surely. The case where $DG = \sigma^2$ a.s. recovers (1.10), and hence occurs if and only if $W^*$ has the $W$-zero bias distribution.

The important message of Lemma 2.1 below is that when $D$ and $G$ in a given Stein or zbest coupling satisfy $DG \geq 0$ then one may use that coupling to produce a new zbest coupling with $DG = \sigma^2$, and therefore one that yields a variable with the $W$-zero bias distribution. In particular, the new zbest coupling is constructed by biasing the original one via the Radon Nikodym factor $DG/E[DG] = DG/\sigma^2$. Further, if the non-negativity condition is not satisfied, one can may nevertheless be able to construct a variable with a distribution close to that of the $W$-zero bias, and achieve a bound to the normal that has an additional term, see Theorem 2.3.

We illustrate the construction of the zero bias distribution as described in Lemma 2.1 in a few simple examples. To start, this construction was previously understood and applied only in the special of the Stein exchangeable pair $(W'', W')$ with distribution $F$ satisfying the linear regression condition (1.6). In this case, identity (1.7) shows that, after trivially renaming the variables there, (1.5) holds with $\mu = 0$ and $G = (W'' - W')/2\lambda$. In this case $DG = (W'' - W')^2/2\lambda$ is non-negative, and non-trivial by (2.3), so that one may construct a new pair $(W^\perp, W^\perp)$ via the Radon Nikodym factor $DG/\sigma^2$, that is, by

$$dF^\perp(W'', W') = \frac{(W'' - W')^2}{2\lambda\sigma^2}dF(W'', W').$$

Lemma 2.1 shows that this construction produces a new zbest coupling that yields a variable with the zero bias distribution.

A rather simple special case is produced by taking a random variable $W$ with mean $\mu = 0$ and noting that $(W, 0, W)$ trivially satisfies (1.11), and that $DG = W^2$ is non-negative. In this case, as is known, and follows from Lemma 2.1 by taking

$$dF^\square(w) = \frac{w^2}{\sigma^2}dF(w)$$

one obtains $W^* = UW^\square$ (2.4) with the $W$-zero bias distribution, where $W^\square$ has distribution (2.4), and is independent of $U$. 

---

Page 4/12
Lemma 2.1. For $W$ having mean $\mu$ and variance $\sigma^2 \in (0, \infty)$, let $(W''', W', G)$ be a zbest coupling on the probability space $(\Omega, \mathcal{F}, P)$, $D = W''' - W'$, and $R$ a P a.s. non-negative random variable satisfying $E_P[R] = 1$ and

$$1(D \neq 0) \leq 1(R \neq 0) \quad \text{a.s. } P.$$  \hspace{1cm} (2.5)

Then $(W''', W', G/R)$ is zbest coupling under the measure

$$dP^\dagger = RdP.$$  \hspace{1cm} (2.6)

If

$$DG \geq 0 \quad \text{and} \quad 1(D \neq 0) \leq 1(G \neq 0) \quad \text{a.s. } P,$$

then one may choose $R = DG/\sigma^2$, and the resulting $X^*$ of (2.1) has the $W$-zero bias distribution.

Proof: As $P^\dagger(R = 0) = 0$ the ratio $G/R$ is well defined under $P^\dagger$, and changing measure we have

$$E_{P^\dagger} \left[ \left( \frac{G}{R} \right) (f(W''') - f(W')) \right] = E_{P^\dagger} \left[ \left( \frac{G}{R} \right) 1(R \neq 0)(f(W''') - f(W')) \right]$$

$$= E_P \left[ \left( \frac{G}{R} \right) R1(R \neq 0)(f(W''') - f(W')) \right] = E_P [ G1(R \neq 0)(f(W''') - f(W')) ]$$

$$= E_P[DG - f(W')] = E_P[(W - \mu)f(W)],$$

where we have used (2.5) for the fourth equality, which, upon noting that $f(W'') - f(W') = 0$ on the event $D = 0$, yields

$$1(R \neq 0)(f(W''') - f(W')) = 1(R \neq 0, D \neq 0)(f(W''') - f(W'))$$

$$= 1(D \neq 0)(f(W''') - f(W')) = f(W''') - f(W').$$

Hence $(W''', W', G/R)$ is a zbest coupling under $P^\dagger$.

Under (2.6), the variable $R = DG/\sigma^2$ is non-negative, integrates to 1 under $P$ by (2.3), and satisfies (2.5), as

$$1(D \neq 0) = 1(D \neq 0)1(G \neq 0) = 1(DG \neq 0).$$

By the first claim, letting $D' = W''' - W'$ and $G' = \sigma^2/D'$, we have $(W''', W', G')$ is a zbest coupling under $P^\dagger$, noting also that $D' \neq 0$ a.s. $P^\dagger$. The last claim now follows from (2.2) noting that $G'D' = \sigma^2$. \hfill \blacksquare

To help recall the construction in Lemma 2.1 via the change of measure to $P^\dagger$, and to better distinguish from Stein couplings that may appear in their proximity, we may now denote a zbest coupling triple by $(W^\dagger, W^\triangledown, G^\dagger)$, and correspondingly let $D^\dagger = W^\dagger - W^\triangledown$.

Remark 2.2. One important instance not previously noted in the literature that the generalization in Lemma 2.1 brings to light begins with the size bias coupling for a non-negative, mean zero, variance $\sigma^2 \in (0, \infty)$ random variable $W$. We recall that $(W'', W', G) = (W^*, W, \mu)$ is a Stein coupling when $W^*$ has the $W$-size biased distribution, characterized by the satisfaction of (1.8). In this case, $DG = \mu(W^* - W)$ is non-negative when the coupling is monotone, that is, when $W^* \geq W$, and by (2.3) $R = \mu(W^* - W)/\sigma^2 = (W^* - W)/E[W^* - W]$ is non-negative and integrates to one. In this case, inequality (2.6) is trivially satisfied, and Lemma 2.1 shows that the linear interpolation $W^* = W^\dagger + (1 - U)W^\triangledown$ by an independent standard uniform variable $U$ of $(W'', W')$ under the change of measure

$$dP^\dagger = \frac{W^\dagger - W^\triangledown}{E[W^\dagger - W^\triangledown]}dP$$  \hspace{1cm} (2.7)

yields $W^*$ with the zero bias distribution.
Enhanced Stein Couplings

For normal approximation, the following theorem yields bounds that imply existing ones in [13] and [5] for zero-bias couplings in both the Wasserstein and Kolmogorov metrics, as in that case the additional term $E[G^1D^1 - \sigma^2]$ vanishes. In addition, (2.9) generalizes the Kolmogorov bound given in Theorem 5.1 of [5] by relaxing the condition that $W^*$ have the $\mathcal{W}$-zero bias distribution.

**Theorem 2.3.** Let $W$ be a random variable with mean zero and variance 1. Then for any coupling of $W$ and a zbest triple $(W^1, W^1, G^1)$, letting $W^* = UW^1 + (1 - U)W^1$, we have

$$d(W, Z) \leq 2E|W^* - W| + \sqrt{2/\pi}E|1 - G^1D^1|,$$

and when $|W^* - W| \leq \delta$ for some $\delta \geq 0$,

$$d_K(W, Z) \leq \left(\frac{1}{\sqrt{2\pi}} + \frac{\sqrt{2\pi}}{4} + 1\right) \delta + E|1 - G^1D^1|$$

where the constant multiplying $\delta$ in the bound is less than 2.03.

**Proof:** For the Wasserstein bound take $h \in \text{Lip}_1$ and let $f_h$ be the unique bounded solution of the Stein equation (1.3). For any coupling of $W$ and $(W^1, W^1, G^1)$, by (1.3) and (2.2)

$$E[h(W) - Nh] = E[f_h^\prime(W) - Wf_h'(W)] = E[f_h^\prime(W) - G^1D^1f_h'(W^*)]$$

$$= E[f_h^\prime(W) - f_h'(W^*) + (1 - G^1D^1)f_h'(W^*)] \leq 2E|W^* - W| + \sqrt{2/\pi}E|1 - G^1D^1|,$$

where we have applied the mean value theorem to obtain the first term, and the bounds $\|f_h''\| \leq 2$ and $\|f_h'\| \leq \sqrt{2/\pi}$ from (2.13) of Lemma 2.4 in [5]. Taking supremum on the left hand side over $h \in \text{Lip}_1$ yields the result.

For the Kolmogorov bound, given an arbitrary $z \in \mathbb{R}$, by Lemma 2.2 in [5], the unique bounded solution $f_z$ of the Stein equation (1.3) with $h(w) = 1_{(-\infty, z]}(w)$ is given by

$$f_z(w) = \sqrt{2\pi}w^{1/2}(\Phi(w)(1 - \Phi(z))1_{w \leq z} + \Phi(z)(1 - \Phi(w))1_{w > z}), \quad (2.10)$$

where $\Phi$ is the cumulative distribution function of the standard normal. We set $f_z^\prime(z) = \lim_{w \to z} f_z^\prime(w)$ so that the first equality in (1.3) holds for all $w \in \mathbb{R}$. We have

$$f_{z-\delta}^\prime(W^*) = 1(W^* \leq z - \delta) - \Phi(z - \delta) + W^*f_{z-\delta}(W^*)$$

$$\leq 1(W \leq z) - \Phi(z - \delta) + W^*f_{z-\delta}(W^*),$$

and so

$$1(W \leq z) - \Phi(z) = (\Phi(z - \delta) - \Phi(z)) + 1(W \leq z) - \Phi(z - \delta)$$

$$\geq -\frac{\delta}{\sqrt{2\pi}} + f_{z-\delta}^\prime(W^*) - W^*f_{z-\delta}(W^*).$$

Taking expectation, applying (2.2) and using that $|f_z^\prime(w)| \leq 1$ for all real $w, z$ via (2.8) of Lemma 2.3 of [5] in the final inequality yields

$$P(W \leq z) - \Phi(z) \geq -\frac{\delta}{\sqrt{2\pi}} + E[f_{z-\delta}^\prime(W^*) - W^*f_{z-\delta}(W^*)]$$

$$= -\frac{\delta}{\sqrt{2\pi}} + E[G^1D^1f_{z-\delta}(W^*) - W^*f_{z-\delta}(W^*) + (1 - G^1D^1)f_{z-\delta}(W^*)]$$

$$\geq -\frac{\delta}{\sqrt{2\pi}} - E|Wf_{z-\delta}(W) - W^*f_{z-\delta}(W^*)| - E|1 - G^1D^1|.$$  (2.11)
For the second term, applying the bound
\[ |(w + u)f_z(w + u) - (w + v)f_z(w + v)| \leq \left( |w| + \sqrt{\frac{2\pi}{4}} \right) (|u| + |v|) \quad \text{for all } u, v, w \in \mathbb{R} \]
from (2.10) of Lemma 2.3 in [5], and letting \( D = W^* - W \), we have
\[
E|f_{z-\delta}(W) - W^*f_{z-\delta}(W^*)| = E|f_{z-\delta}(W) - (W + D)f_{z-\delta}(W + D)| \leq E \left( |W| + \sqrt{\frac{2\pi}{4}} |D| \right) \leq \left( 1 + \sqrt{\frac{2\pi}{4}} \right) \delta,
\]
using \( E|W| \leq \sqrt{EW^2} = 1 \). Substituting into (2.11) produces a lower bound on \( P(W \leq z) - \Phi(z) \). Proving an analogous upper bound by the same reasoning, and then taking supremum over \( z \in \mathbb{R} \) yields the claim (2.9).

3 The Lightbulb Process

The lightbulb problem was first considered in [27] as a model for the behavior of skin receptors subject to a medication released by dermal patches, and was then subsequently studied in [17]. In the model, in a sequence of \( n \) stages, \( n \) skin receptors, which here we will imagine as lightbulbs, are toggled from one state to the other upon absorbing a pharmaceutical. Initially all \( n \) lightbulbs are turned off. Letting \( [n] = \{1, \ldots, n\} \), at stage \( r \in [n] \) a set of \( r \) lightbulbs, selected uniformly at random from all subsets of \( [n] \) of size \( r \), and independently of previous stages, have their state toggled.

We study the random variable \( Y \) that counts the number of lightbulbs that are turned on after stage \( n \) is complete. This framework can be considered more generally, where in stage \( r \) the number of toggled lightbulbs is \( s_r \in [n] \), see [27] and [17].

More formally consider \( X \in \{0, 1\}^{n \times n} \), a matrix of Bernoulli variables, called here a configuration, and whose components we refer to as toggle variables. The initial state of the bulbs is given deterministically, setting all bulbs off. For stages \( r \in [n] \) the components of \( X \) have the interpretation that
\[
X_{rj} = \begin{cases} 
1 & \text{if the status of bulb } j \text{ is changed at stage } r, \\
0 & \text{otherwise.}
\end{cases}
\]
As exactly \( r \) of the \( n \) bulbs are chosen uniformly to be toggled at stage \( r \), taking
\[
E = \left\{ e \in \{0, 1\}^{n \times n} : \sum_{j=1}^{n} e_{rj} = r, r = 1, \ldots, n \right\},
\]
and the toggles in each stage \( r \in [n] \) independent, the distribution of \( X \) is given by
\[
P(X = e) = \left\{ \begin{array}{ll}
\prod_{r=1}^{n} \binom{n}{r}^{-1} & e \in E \\
0 & \text{otherwise.}
\end{array} \right. \tag{3.1}
\]
The toggle variables at stage \( r \) that form the vector \((X_r, \ldots, X_n)\) are clearly exchangeable and the marginal distribution of the components \( X_{rj} \) are Bernoulli with success probability \( r/n \). For bulbs \( i \in [n] \), the variables
\[
X_i = \left( \sum_{r=1}^{n} X_{ri} \right) \mod 2 \quad \text{and} \quad Y = \sum_{i=1}^{n} X_i \tag{3.2}
\]
Taking marginals, one immediately finds that

\[ Y \quad \text{and} \quad X \]

quantity result in a configuration then all quantities are again zero as interchanging the toggles in the middle stage will result in a configuration \( Y'' = Y \) if \( X_I = 1 \) and otherwise letting \( Y'' \) be the sum of indicators having distribution

\[
P(X_i = e_1, \ldots, X_n = e_n) = P(X_1 = e_1, \ldots, X_n = e_n | X_I = 1) \quad (3.3)
\]

for \( i = I \). We now show how to apply Lemma 2.1 to create a zbest coupling from this size bias coupling.

Given a configuration \( X \in \mathcal{E} \), for a stage \( s \in [n] \) and two indices \( i, j \in [n] \), let \( X_{s,i\leftrightarrow j} \) be the configuration with components

\[
X_{s,i\leftrightarrow j} = \begin{cases} 
X_{rk} & r \neq s \\
X_{rk} & r = s, k \not\in \{i, j\} \\
X_{rj} & r = s, k = i \\
X_{ri} & r = s, k = j,
\end{cases}
\]

that is, the new configuration is the same as \( X \), but with the toggle variables \( X_{si} \) and \( X_{sj} \) interchanged.

Taking \( n \geq 4 \) to be even for simplicity, a size biased coupling \((Y'', Y)\) was constructed in \([17]\) as follows; the odd case can be handled using randomization as in \([17]\). First, sample \( X \) according to \((3.1)\). Given \( X \), sample \( I \) uniformly from \([n]\), and given \( X \) and \( I \), let \( J \) be an index chosen uniformly at random over the set of \( n/2 \) indices \( \{j \in [n] : X_{n/2,j} \neq X_{n/2,\frac{n}{2}}\} \). Finally, let \( X'' = X \) if \( X_I = 1 \) and \( X''=X_{n/2,\frac{n}{2}} \) otherwise. Hence, with \( e_i = (\sum_n e_{ri}) \mod 2 \) as in \((3.2)\), and \( e, e'' \in \mathcal{E} \), we have

\[
P(X = e, X'' = e'', I = i, J = j) = \frac{21(e_{n/2,i} \neq e_{n/2,j})}{n^2 \prod_{k=1}^{n} \binom{n}{2}} \left(1(e_i = 1, e'' = e) + 1(e_i = 0, e'' = e''_{n/2,\frac{n}{2}})\right). \quad (3.4)
\]

Taking marginals, one immediately finds that

\[
P(X = e, I = i, J = j) = \eta 1(e_{n/2,i} \neq e_{n/2,j}) \quad \text{where} \quad \eta = \frac{2}{n^2 \prod_{k=1}^{n} \binom{n}{2}} \quad (3.5)
\]

In \([17]\), via a verification of \((3.3)\), it was shown that \( Y'' \), the number of lightbulbs that are on in the terminal stage in configuration \( X'' \), has the \( Y \) size bias distribution and satisfies

\[
Y'' - Y = 21\{X_I = 0, X_J = 0\} = 21\{Y'' \neq Y\}. \quad (3.6)
\]

Referring the reader to \([17]\) for a proof of the size property, we show \((3.6)\). Indeed, if \( X_I = 1 \) then \( Y'' = Y \) and the three quantities above are all zero. If \( X_I = 0 \) and \( X_J = 1 \) then all quantities are again zero as interchanging the toggles in the middle stage will result in a configuration \( X'' \) in which \( X'_I = 1, X'_J = 0 \), thus leaving \( Y'' \) with the same value as \( Y \). In the remaining case \( X_I = 0, X_J = 0 \) and all quantities above equal 2, as the exchange toggles the final state of both bulbs \( I \) and \( J \), turning both from off to on, increasing \( Y \) by 2.
Identity (**3.6**) demonstrates in particular that the size bias coupling considered here is monotone, and thus Remark (**2.2**) is in force. To continue, one needs to construct the biased version of the original pair according to $P^\dagger$ in (**2.7**). Starting with (**3.4**) and making use of the first equality in (**3.6**), we see that the $P^\dagger$ distribution is given by

$$
P^\dagger(X = e, X'' = e'', I = i, J = j) = \frac{y'' - y}{E[Y'' - Y]} P(X = e, X'' = e'', I = i, J = j)
$$

$$
= \frac{1(e_i = 0, e_j = 0)}{P(Y'' \neq Y)} P(X = e, X'' = e'', I = i, J = j)
$$

$$
= \frac{\eta}{P(Y'' \neq Y)} 1(e_{n/2, i} \neq e_{n/2, j}, e_i = 0, e_j = 0, e'' = e^{n/2,i+j}),
$$

where in the last equality we have applied (**3.4**) and the definition of $\eta$ in (**3.5**). Integrating out the distribution of $X''$, we find the marginal of $X$ under $P^\dagger$ satisfies

$$
P^\dagger(X = e, I = i, J = j) = \frac{\eta}{P(Y'' \neq Y)} 1(e_{n/2, i} \neq e_{n/2, j}, e_i = 0, e_j = 0).
$$

**Lemma 3.1.** There exists a coupling of $(X', X'')$ and $(X^\dagger, X^\ddagger)$, having distributions specified by (**3.4**) and (**3.7**), respectively, that satisfies

$$
|Y^\dagger - Y''| \leq 2 \quad \text{and} \quad Y^\dagger = Y^\ddagger + 2.
$$

**Proof:** To achieve the required construction, first sample $X, I, J$ from $P$ as in (**3.5**). Next, let $S$ be any random variable with support in $[n - 1] \setminus \{n/2\}$, independent of $X, I, J$. Sample indices $K$ and $L$ that have marginal conditional laws $L(\cdot | \cdot)$ specified by

$$
L(K | X, I, J, S) \sim \mathcal{U}\{k \notin \{I, J\} : X_{S,k} \neq X_{S,I}\}
$$

and

$$
L(L | X, I, J, S) \sim \mathcal{U}\{l \notin \{I, J\} : X_{S,l} \neq X_{S,J}\}.
$$

Now for $(a, b) \in \{0, 1\}^2$, suppressing the dependence of $\phi_{ab}$ on $I, J, K, L$ and $S$, let

$$
X^\dagger = \phi_{X, I, J}(X) \quad \text{where} \quad \phi_{ab}(e) = \begin{cases} 
\phi_{00}(e) := e & \text{if } (a, b) = (0, 0) \\
\phi_{01}(e) := e_{S, J+L} & \text{if } (a, b) = (0, 1) \\
\phi_{10}(e) := e_{S, I+K} & \text{if } (a, b) = (1, 0) \\
\phi_{11}(e) := e^{n/2, I+J} & \text{if } (a, b) = (1, 1).
\end{cases}
$$

One easily verifies that $P(X^\dagger \in \mathcal{E}) = 1$ and that each $\phi_{ab}$ is an involution mapping between $\{e \in \mathcal{E} : e_i = a, e_j = b\}$ and $\{e \in \mathcal{E} : e_i = 0, e_j = 0\}$. Moreover, for any $f \in \mathcal{E}$,

$$
P(X^\dagger = f, I = i, J = j) = P(\phi_{X, I, J}(X) = f, I = i, J = j)
$$

$$
= \sum_{(a, b) \in \{0, 1\}^2} P(\phi_{ab}(X) = f, I = i, J = j, X_i = a, X_j = b).
$$

Note that on the event $\{I = i, J = j\}$ we have $X_{n/2, i} \neq X_{n/2, j}$ and the probability in the $a, b$th summand is zero unless $(\phi_{ab}(f))_{n/2, i} \neq (\phi_{ab}(f))_{n/2, j}$, which is equivalent to the condition that $f_{n/2, i} \neq f_{n/2, j}$. Similarly, for this probability to be non-zero we must have that $(\phi_{ab}(f))_{i} = a, (\phi_{ab}(f))_{j} = b$, which implies that $f_i = 0, f_j = 0$. For $f \in \mathcal{E} : f_{n/2, i} \neq f_{n/2, j}, f_i = 0, f_j = 0$ the restriction in the summand that $X_i = a, X_j = b$ is redundant, and as $\phi_{ab}(f) \in \mathcal{E}$ and the probability $P(X = e, I = i, J = j)$ is constant over its support by (**3.8**), we obtain

$$
P(X^\dagger = f, I = i, J = j) \propto 1(f_{n/2, i} \neq f_{n/2, j}, f_i = 0, f_j = 0),
$$

where
Theorem 3.2. The claims now respectively follow from the two final inequalities, and (2.8) and (2.9), where the first equality follows from a simple symmetry argument, and the next from valid.

\[ \bullet \]

The mean \( \mu \) and variance \( \sigma^2 \) of \( Y \) are given by

\[
\mu = n/2 \quad \text{and} \quad \sigma^2 = \frac{n}{4} \left(1 + (n-1)\lambda_n\right) \quad \text{with} \quad \lambda_n = \prod_{s=1}^{n} \left(1 - \frac{4s(n-s)}{n(n-1)}\right),
\]

where the first equality follows from a simple symmetry argument, and the next from Section 2.3 of [27], and as given in (3), (4) and (5) of [17], where \( \lambda_n \) there is denoted \( \lambda_{n,2,n} \). Note that when \( n \) is even, as the terms in the product for \( s \) and \( n-s \) are equal,

\[
\lambda_n = \left(1 - \frac{4(n/2)^2}{n(n-1)}\right) \prod_{s=1}^{n/2-1} \left(1 - \frac{4s(n-s)}{n(n-1)}\right)^2 = -\frac{1}{n-1} \prod_{s=1}^{n/2-1} \left(1 - \frac{4s(n-s)}{n(n-1)}\right)^2 \leq 0,
\]

and hence \( \sigma^2 \leq n/4 \). Applying the reasoning that proves (1.6) of [8] demonstrates that the order \( 1/\sigma \) in the Kolmogorov bound in [17], and (3.10), is unimprovable.

**Theorem 3.2.** Let \( Y \) be the number of lightbulbs switched on in the final stage of the lightbulb process with \( n \geq 4 \) even stages. Then, with \( \mu = E[Y], \sigma^2 = \text{Var}(Y) \) and \( W = (Y - \mu)/\sigma \), with \( Z \sim \mathcal{N}(0,1), \)

\[
d(W, Z) \leq 6/\sigma \quad \text{and} \quad d_K(W, Z) \leq 8.12/\sigma.
\]

**Proof:** Consider the coupling given in Lemma 3.1. Letting \( U \sim U[0,1], \) independent of \( Y, Y^\dagger, Y^\ddagger, \) and \( W^\dagger = (Y^\dagger - \mu)/\sigma \) and \( W = (Y - \mu)/\sigma, \) by Remark 2.2 and scaling,

\[
W^\ast = UW^\dagger + (1 - U)W^\ddagger
\]

has the \( W^{-}\)zero bias distribution, and hence \( D\dagger G\dagger = 1. \) Using (3.9) of Lemma 3.1

\[
|W^\ast - W| = |U(Y^\dagger + 2) + (1 - U)Y^\ddagger - Y|/\sigma = |Y^\dagger + Y^\ddagger - 2U|/\sigma \leq (2 + 2U)/\sigma \leq 4/\sigma.
\]

The claims now respectively follow from the two final inequalities, and (2.8) and (2.9), respectively.

Now with

\[
\Delta_0 = \frac{1}{2\sqrt{n}} + \frac{1}{2n} + \frac{1}{3}n^{-n/2},
\]

and \( B_n \) denoting the constant in [17] that for \( n \geq 6 \) replaces 8.12 in (3.10), over that range, using \( \sigma^2 \leq n/4 = \sigma^2 \), we have

\[
\sigma B_n = \frac{n}{2\sigma} \Delta_0 + 1.64 \frac{n}{\sigma} + 2 \geq \frac{n}{2\sigma} \frac{1}{2\sqrt{n}} + 1.64 \frac{n}{\sigma} + 2 = \frac{1}{2} + 4(1.64) + 2 = 9.06.
\]

Hence the bound produced here is superior for all values over which the former was valid.

**References**

[1] Sourav Chatterjee, *Stein’s method for concentration inequalities*, arXiv preprint math/0604352 (2006).

[2] Sourav Chatterjee and Partha S Dey, *Applications of Stein’s method for concentration inequalities*, The Annals of Probability 38 (2010), no. 6, 2443–2485.
Enhanced Stein Couplings

[3] Sourav Chatterjee, Jason Fulman, and Adrian Röllin, *Exponential approximation by Stein’s method and spectral graph theory*, ALEA Lat. Am. J. Probab. Math. Stat 8 (2011), no. 1, 197–223.

[4] Louis HY Chen, *Poisson approximation for dependent trials*, The Annals of Probability 3 (1975), no. 3, 534–545.

[5] Louis HY Chen, Larry Goldstein, and Qi-Man Shao, *Normal approximation by Stein’s method*, Springer Science & Business Media, 2010.

[6] Louis HY Chen and Adrian Röllin, *Stein couplings for normal approximation*, arXiv preprint arXiv:1003.6039 (2010).

[7] Christian Döbler, *New Berry-Esseen and Wasserstein bounds in the CLT for non-randomly centered random sums by probabilistic methods*, arXiv preprint arXiv:1504.05938 (2015).

[8] Gunnar Englund, *A remainder term estimate for the normal approximation in classical occupancy*, The Annals of Probability (1981), 684–692.

[9] Murat A Erdogdu, Lee H Dicker, and Mohsen Bayati, *Scaled least squares estimator for GLMs in large-scale problems*, Advances in Neural Information Processing Systems 29 (2016).

[10] Max Fathi, Larry Goldstein, Gesine Reinert, and Adrien Saumard, *Relaxing the Gaussian assumption in shrinkage and SURE in high dimension*, arXiv preprint arXiv:2004.01378 (2020).

[11] Robert E Gaunt, *Stein factors for variance-gamma approximation in the Wasserstein and Kolmogorov distances*, arXiv preprint arXiv:2008.06088 (2020).

[12] Subhankar Ghosh and Larry Goldstein, *Concentration of measures via size-biased couplings*, Probability Theory and Related fields 149 (2011), no. 1, 271–278.

[13] Larry Goldstein, *Normal approximation for hierarchical structures*, The Annals of Applied Probability 14 (2004), no. 4, 1950–1969.

[14] ______, *$L^1$ bounds in normal approximation*, The Annals of Probability 35 (2007), no. 5, 1888–1930.

[15] ______, *Non-asymptotic distributional bounds for the Dickman approximation of the running time of the Quickselect algorithm*, Electronic Journal of Probability 23 (2018), 1–13.

[16] Larry Goldstein and Gesine Reinert, *Stein’s method and the zero bias transformation with application to simple random sampling*, The Annals of Applied Probability 7 (1997), no. 4, 935–952.

[17] Larry Goldstein and Haimeng Zhang, *A Berry-Esseen bound for the lightbulb process*, Advances in Applied Probability 43 (2011), no. 3, 875–898.

[18] Tobias Johnson and Erol Peköz, *Concentration inequalities from monotone couplings for graphs, walks, trees and branching processes*, arXiv preprint arXiv:2108.02101 (2021).

[19] Günter Last, Giovanni Peccati, and Matthias Schulte, *Normal approximation on Poisson spaces: Mehler’s formula, second order Poincaré inequalities and stabilization*, Probability Theory and Related Fields 165 (2016), no. 3, 667–723.

[20] Qiang Liu, Jason Lee, and Michael Jordan, *A kernelized Stein discrepancy for goodness-of-fit tests*, International conference on machine learning, PMLR, 2016, pp. 276–284.

[21] Qiang Liu and Dilin Wang, *Stein variational gradient descent: a general purpose Bayesian inference algorithm*, Advances in Neural Information Processing Systems 29 (2016).

[22] Ho Ming Luk, *Stein’s method for the gamma distribution and related statistical applications*, Ph.D. thesis, University of Southern California, 1994.

[23] Domenico Marinucci and Maurizia Rossi, *Stein–Malliavin approximations for nonlinear functionals of random eigenfunctions on $S^d$*, Journal of Functional Analysis 268 (2015), no. 8, 2379–2420.

[24] Ivan Nourdin and Giovanni Peccati, *Normal approximations with Malliavin calculus: from Stein’s method to universality.*

[25] Erol A Peköz, *Stein’s method for geometric approximation*, Journal of Applied Probability 33 (1996), no. 3, 707–713.
[26] Erol A Peköz, Adrian Röllin, and Nathan Ross, *Degree asymptotics with rates for preferential attachment random graphs*, The Annals of Applied Probability 23 (2013), no. 3, 1188–1218.

[27] C Radhakrishna Rao, M Bhaskara Rao, and Haimeng Zhang, *One bulb? Two bulbs? How many bulbs light up?—a discrete probability problem involving dermal patches*, Sankhyä: The Indian Journal of Statistics (2007), 137–161.

[28] Nathan Ross, *Fundamentals of Stein’s method*, Probability Surveys 8 (2011), 210–293.

[29] Charles Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, Proceedings of the sixth Berkeley symposium on mathematical statistics and probability, volume 2: Probability theory, University of California Press, 1972, pp. 583–602.