A Berry Esseen Theorem for the Lightbulb Process

Larry Goldstein and Haimeng Zhang
University of Southern California and Mississippi State University

January 4, 2010

Abstract

In the so called lightbulb process, on days \( r = 1, \ldots, n \), out of \( n \) lightbulbs, all initially off, exactly \( r \) bulbs, selected uniformly and independent of the past, have their status changed from off to on, or vice versa. With \( X \) the number of bulbs on at the terminal time \( n \), an even integer, and \( \mu = n/2, \sigma^2 = \text{Var}(X) \), we have

\[
\sup_{z \in \mathbb{R}} \left| P \left( \frac{X - \mu}{\sigma} \right) - P(Z \leq z) \right| \leq \frac{n}{2\sigma^2} \Delta_0 + 1.64 \frac{n}{\sigma^3} + \frac{2}{\sigma}
\]

where

\[
\Delta_0 \leq \frac{1}{2\sqrt{n}} + \frac{1}{2n} + e^{-n/4} \quad \text{for} \quad n \geq 4,
\]

yielding a bound of order \( O(n^{-1/2}) \) as \( n \to \infty \). A similar, though slightly larger bound holds for \( n \) odd.

The results are shown using a version of Stein’s method for bounded, monotone size bias couplings. The argument for even \( n \) depends on the construction of a variable \( X^s \) on the same space as \( X \) which has the \( X \) size bias distribution, that is, that satisfies

\[
EXg(X) = \mu Eg(X^s)
\]

for all bounded continuous \( g \), and for which there exists a \( B \geq 0 \), in this case \( B = 2 \), such that \( X \leq X^s \leq X + B \) almost surely. The argument for \( n \) odd is similar to that for \( n \) even, but one first couples \( X \) closely to \( V \), a symmetrized version of \( X \), for which a size bias coupling of \( V \) to \( V^s \) can proceed as in the even case.

1 Introduction

The problem we consider here arises from a study in the pharmaceutical industry on the effects of dermal patches designed to activate targeted receptors. An active receptor will become inactive, and an inactive one active, if it receives a dose of medicine released from the dermal patch. Let the number of receptors, all initially inactive, be denoted by \( n \). On study day 1, 2, \ldots, some number \( s \) of \( n \) randomly selected receptors will each receive one dose of medicine, changing their statuses between the inactive and active states. We adopt the following, somewhat more colorful, though equivalent, 'lightbulb process' formulation. Consider \( n \) toggle switches, each being connected to a lightbulb. Pressing the toggle switch connected to a bulb changes its status from off to on and vice versa. The problem of determining the properties of \( X \), the number of light bulbs on at the end of day \( n \), was first considered in [6] for the case where on each day \( s = 1, \ldots, n \), exactly \( s \) of the \( n \) switches are randomly pressed.

Consider the lightbulb process with some number \( k \) of stages where \( s_r \in \{0, \ldots, n\} \) lightbulbs are toggled in stage \( r \), for \( r = 1, \ldots, k \), and let \( s = (s_1, \ldots, s_k) \). To track the status of some subset of \( b \) of the \( n \) bulbs, we define

\[
\lambda_{n,b,s} = \sum_{t=0}^{b} \binom{b}{t} (-2)^t \binom{s}{t} \binom{n}{t} \quad \text{and} \quad \lambda_{n,b,s} = \prod_{r=1}^{k} \lambda_{n,b,s_r},
\]
where \((n)_k = n(n-1) \cdots (n-k+1)\) denotes the falling factorial, and the empty product is 1. In particular
\[
\lambda_{n,1,s} = 1 - \frac{2s}{n} \quad \text{and} \quad \lambda_{n,2,s} = 1 - \frac{4s}{n} + \frac{4s(s-1)}{n(n-1)} \quad \text{for} \ s = 1, \ldots, n.
\]

Generalizing the results in [6], writing \(X_s\) for the number of bulbs on at the terminal time when applying \(s_r\) switches in stages \(r = 1, \ldots, n\), the martingale method in Proposition 4 of [10] shows that if the chain is initialized with all bulbs off, then
\[
EX_s = \frac{n}{2} (1 - \lambda_{n,1,s}) \quad \text{and} \quad \text{Var}(X_s) = \frac{n}{4} (1 - \lambda_{n,2,s}) + \frac{n^2}{4} (\lambda_{n,2,s} - \lambda_{n,1,s}^2).
\]

In particular, the mean \(\mu = EX\) and variance of \(\sigma^2 = \text{Var}(X)\) of \(X\) as considered in [6] are given by (2) with \(s = [n]\) where \([n] = \{1, \ldots, n\}\). Other results in [6] include recursions for determining the exact finite sample distribution of \(X\). Though approximations to the distribution of \(X\), including by the normal, were also considered, the quality of such approximations, and the asymptotic normality of \(X\) was left open.

The following theorem settles the question of the asymptotic distribution of \(X\) by providing a bound to the normal which holds for all finite \(n\), and which tends to zero at the rate \(n^{-1/2}\) as \(n\) tends to infinity. For \(n = 2m + 1\) an odd number, let

\[
s_{m,m} = (1, \ldots, m-1, m, m, m+2, \ldots, n), \quad s_{m+1,m+1} = (1, \ldots, m-1, m+1, m+1, m+2, \ldots, n)
\]

and

\[
\lambda_{n,b} = \frac{1}{2} \left( \lambda_{n,b,s_{m,m}} + \lambda_{n,b,s_{m+1,m+1}} \right).
\]

We proceed in the odd case by coupling \(X\) to a more symmetric random variable \(V\) with mean and variance given respectively by
\[
\mu_V = \frac{n}{2} (1 - \lambda_{n,1}) \quad \text{and} \quad \sigma_V^2 = \frac{n^2}{16} (\lambda_{n,1,s_{m,m}} - \lambda_{n,1,s_{m+1,m+1}})^2 + \frac{n}{4} (1 - \lambda_{n,2}) + \frac{n^2}{4} \left( \lambda_{n,2} - \frac{\lambda_{n,1,s_{m,m}}^2 + \lambda_{n,1,s_{m+1,m+1}}^2}{2} \right).
\]

**Theorem 1.1** Let \(X\) be the number of bulbs on at the terminal time \(n\) in the lightbulb process, applying \(s_r = r\) switches in stages \(r = 0, 1, \ldots, n\). For all \(n\) even

\[
\sup_{z \in \mathbb{R}} \left| P \left( \frac{X - n/2}{\sigma} \leq z \right) - P(Z \leq z) \right| \leq \frac{n}{2\sigma^2} \Delta_0 + 1.64 \frac{n}{\sigma^3} + \frac{2}{\sigma},
\]

where \(\sigma^2 = \text{Var}(X)\) is given by (2) and

\[
\Delta_0 \leq \frac{1}{2\sqrt{n}} + \frac{1}{2n} + e^{-n/2}, \quad \text{for} \ n \geq 6.
\]

For all \(n = 2m + 1\) odd,

\[
\sup_{z \in \mathbb{R}} \left| P \left( \frac{X - \mu_V}{\sigma_V} \leq z \right) - P(Z \leq z) \right| \leq \frac{\mu_V}{\sigma_V^2} \Delta_1 + 3.28 \frac{\mu_V}{\sigma_V} + \frac{1}{\sigma_V} \left( 2 + \frac{1}{\sqrt{2\pi}} \right),
\]

where \(\mu_V\) and \(\sigma_V^2\) are given in (4) and

\[
\Delta_1 \leq \frac{1}{2\sqrt{n}} + \frac{\sqrt{5}}{4n} + \frac{1}{n\sqrt{2(n-1)}} + \frac{\sqrt{7}}{n} e^{-n/4} + e^{-n/2}, \quad \text{for} \ n \geq 7.
\]
In the even case, as \( \lambda_{n,1,[n]} \) and \( \lambda_{n,2,[n]} \) decay exponentially fast to zero, the variance \( \sigma^2 \) is of order \( n \) and the bound (5), therefore, of order \( 1/\sqrt{n} \); analogous remarks hold for the case where \( n \) is odd.

We now more formally describe the lightbulb process. With \( n \in \mathbb{N} \) fixed and \( s = (s_1, \ldots, s_n) \) for \( s_r \in \{0, \ldots, n\} \) with \( r = 1, \ldots, n \), we will let \( X_s = \{X_{rk} : r = 0, 1, \ldots, n, k = 1, \ldots, n\} \) denote a collection of Bernoulli variables, and will write \( X \) for \( X_s \) when \( s = (1, \ldots, n) \). For \( r \geq 1 \) these ‘switch variables’ have the interpretation that

\[
X_{rk} = \begin{cases} 
1 & \text{if the status of bulb } k \text{ is changed at stage } r, \\
0 & \text{otherwise.}
\end{cases}
\]

The initial state of the bulbs is given deterministically by \( \{X_{0k}, k = 1, \ldots, n\} \), which will be taken to be state zero, that is, all bulbs off, unless specifically stated otherwise; in fact, non-zero initial conditions are considered in Lemma 5.2. At stage \( r \), \( s_r \) of the \( n \) bulbs are chosen uniformly to have their status changed, and the stages are independent of each other. Hence, with \( e_1, \ldots, e_n \in \{0,1\} \), the joint distribution of \( X_{r1}, \ldots, X_{rn} \) is given by

\[
P(X_{r1} = e_1, \ldots, X_{rn} = e_n) = \left\{ \begin{array}{cl} \binom{n}{s_r}^{-1} & \text{if } e_1 + \cdots + e_n = s_r, \\ 0 & \text{otherwise,} \end{array} \right. \quad (7)
\]

with the collections \( \{X_{r1}, \ldots, X_{rn}\} \) independent for \( r = 1, \ldots, n \).

Clearly, for each stage \( r \), the variables \( (X_{r1}, \ldots, X_{rn}) \) are exchangeable, and the marginal distribution of \( X_{rk} \) is given by

\[
P(X_{rk} = 1) = \frac{s_r}{n} \quad \text{and} \quad P(X_{rk} = 0) = 1 - \frac{s_r}{n} \quad \text{for all } r, k = 1, \ldots, n.
\]

For \( r, k = 1, \ldots, n \), the quantity \( (\sum_{i=1}^{r} X_{ik}) \mod 2 \) is the indicator that bulb \( k \) is on at time \( r \), so letting

\[
X_k = \left(\sum_{r=0}^{n} X_{rk}\right) \mod 2 \quad \text{and} \quad X_s = \sum_{k=1}^{n} X_k,
\]

the variable \( X_k \) is the indicator that bulb \( k \) is on at the terminal time, and the random variable \( X_s \) is the number of bulbs on at the terminal time.

The lightbulb process, where the \( n \) individual states evolve according to the same marginal Markov chain, is a special case of a class of multivariate chains studied in [10]. As shown in [10], such chains admit explicit full spectral decompositions, and in particular, the transition matrices for each stage of the lightbulb process can be simultaneously diagonalized by a Hadamard matrix. These properties were, in fact, put to use in [6] in particular for the calculation of the moments needed to compute the mean and variance of \( X \) when \( s_r = r, r = 0, \ldots, n \). Here we put these same properties to somewhat more arduous work, the calculation of moments of fourth order.

That no higher order moments are required for the proof of a finite sample bound holding for all \( n \) is one distinct advantage of the technique we apply here, Stein’s method for the normal distribution, brought to life in the seminal monograph [9]. By contrast, the method of moments requires the calculation and appropriate convergence of moments of all orders, and obtains only convergence in distribution. Stein’s method for the normal is based on the characterization of the normal distribution in [8], which states that \( Z \sim \mathcal{N}(0,1) \), that is, that \( Z \) is standard normal, if and only if

\[
E[Zg(Z)] = E[g'(Z)] \quad (9)
\]

for all absolutely continuous functions \( g \) for which these expectations exist. The idea behind Stein’s method is that if a mean zero, variance 1 random variable \( W \) is close in distribution to \( Z \), then \( W \) will satisfy (9) approximately. Hence, to gauge the proximity of \( W \) to \( Z \) for a given test function \( h \), one can evaluate the difference \( Eh(W) - Nh \), where \( Nh = Eh(Z) \), by solving the Stein equation

\[
f'(w) - wf(w) = h(w) - Nh
\]
for $f$ and evaluating $E[f'(W) - W f(W)]$ instead.

A priori it may appear that an evaluation of $E[f'(W) - W f(W)]$ would more difficult than for $E h(W - \sqrt{Y} H)$. However, the former may be handled though the use of couplings. Here we consider size bias couplings in particular. Given a nonnegative random variable $Y$ with positive finite mean $\mu = E Y$, we say $Y^s$ has the $Y$ size biased distribution if $P[Y^s \in dy] = (y/\mu)P[Y \in dy]$, or more formally, if

$$E[Y g(Y)] = \mu E[g(Y^s)] \quad \text{for bounded continuous functions } g.$$  \hspace{1cm} (10)

The use of size-biased couplings in Stein’s method was introduced in [1], where it was used to develop bounds of order $\sigma^{-1/2}$ for the normal approximation to the number of local maxima $Y$ of a random function on a graph, where $\sigma^2 = \text{Var}(Y)$. In [4] the method was extended to multivariate normal approximations, and the rate was improved to $\sigma^{-1}$, for the expectation of smooth functions of a vector $Y$ recording the number of edges with certain fixed degrees in a random graph. In [3] the method was used to give bounds in the Kolmogorov distance of order $\sigma^{-1}$ for various functions on graphs and permutations, and in [5] for two problems in the theory of coverage processes, with bounds of this same order.

Here we prove and apply Theorem 2.1, which requires that $Y$ and a random variable $Y^s$, having the $Y$ size-biased distribution, be constructed on a common space such that for some $B \geq 0$,

$$Y \leq Y^s \leq Y + B$$

with probability one, that is, the coupling must be monotone, and bounded. Loosely speaking, Theorem 2.1 says that given any such coupling of $Y$ and $Y^s$ on a common space, an upper bound on the Kolmogorov distance between the distribution of $Y$ and the normal can be computed in terms of $B$ and the quantity

$$\Delta = \sqrt{\text{Var}(E[Y^s - Y|Y])}.$$  \hspace{1cm} (11)

Theorem 2.1 is based on a concentration type inequality, provided in Lemma 2.1.

For the lightbulb process, a size biased coupling of $X$ to $X^s$ is achieved in the even case by the construction, for each $i = 1, \ldots, n$, of a collection $X^i$ from a given $X$ as follows. If $X_i = 1$, that is, if bulb $i$ is on at the terminal times, we set $X^i = X$. Otherwise, let $J$ be uniformly chosen from all $j$ for which $X_{n/2,j} = 1 - X_{n/2,i}$ and let $X^i$ be the same as $X$ but that the values of $X_{n/2,i}$ and $X_{n/2,J}$ are interchanged. Let $X^i$ be the number of bulbs on at the terminal time when applying the switch variables $X^i$. Then, with $I$ uniformly chosen from $1, \ldots, n$, the variable $X^s = X^I$ has the $X$ size bias distribution, essentially due to the fact, shown in Lemma 3.1, that

$$\mathcal{L}(X^i) = \mathcal{L}(X|X_i = 1).$$

Due to the parity issue, to handle the odd case, say $n = 2m + 1$, we first construct a coupling of $X$ to a more symmetric variable $V$. In particular, $V$ is constructed from the same switch variables as $X$, but that with equal probability either one additional switch variable is applied in stage $m$, or one fewer in stage $m + 1$. A size biased coupling of $V$ to $V^s$ can be achieved as in the even case, thus yielding a bound to the normal for $X$.

In Section 2 we present Theorem 2.1, which gives a bound to the normal when a bounded, monotone size biased coupling can be constructed for a given $X$. Our coupling constructions and proof of the bound for the even case of the lightbulb process are given in Section 3. Calculations of the bounds on the variance $\Delta$ in (11) require estimates on $\lambda_{n,b,s}$ in (1). These estimates are based on the work of [10] which yields the spectral decomposition of the underlying transition matrix. The estimates required, for both the even and odd cases, are given in the Appendix. Symmetrization, that is, the construction of $V$ from $X$, coupling constructions for $V$, and the proof of the bound in the odd case are given in Section 4.

2 Bounded Monotone Couplings

Theorem 2.1 for bounded monotone size bias couplings depends on the following lemma, which is in some sense the size bias version of Lemma 2.1 of [7]. With $Y$ having mean $\mu$ and variance $\sigma^2$, both finite and
positive, with some slight abuse of notation in the definition of $W^*$, we set

$$W = \frac{Y - \mu}{\sigma} \quad \text{and} \quad W^* = \frac{Y^* - \mu}{\sigma}. \quad (12)$$

**Lemma 2.1** Let $Y$ be a nonnegative random variable with mean $\mu$ and variance $\sigma^2$, both finite and positive, and let $Y^*$ be given on the same space as $Y$, having the $Y$ size biased distribution and satisfying $Y^* \geq Y$. Then with $W$ and $W^*$ given in (12), for any $z \in \mathbb{R}$ and $a > 0$,

$$\frac{\mu}{\sigma} E(W^* - W) 1_{\{W^* - W \leq a\}} 1_{\{z \leq W \leq z + a\}} \leq a.$$

**Proof.** For fixed $z \in \mathbb{R}$ let

$$f(w) = \begin{cases} -a & w \leq z \\ w - z - a & z \leq w \leq z + 2a \\ a & w \geq z + 2a. \end{cases}$$

Then, by (12) and (10)

$$a \geq E(Wf(W)) = \frac{1}{\sigma} E(Y - \mu) f\left(\frac{Y - \mu}{\sigma}\right) = \frac{\mu}{\sigma} E(f(W^*) - f(W)) = \frac{\mu}{\sigma} E f(W + t) dt \geq \frac{\mu}{\sigma} E \int_0^{W^* - W} f'(W + t) dt.$$

Noting that $f'(W + t) = 1_{\{z \leq W + t \leq z + 2a\}}$, so that

$$1_{\{0 \leq t \leq a\}} 1_{\{z \leq W \leq z + a\}} f'(W + t) = 1_{\{0 \leq t \leq a\}} 1_{\{z \leq W \leq z + a\}},$$

we obtain

$$a \geq \frac{\mu}{\sigma} E \int_0^{W^* - W} 1_{\{0 \leq t \leq a\}} 1_{\{z \leq W \leq z + a\}} dt = \frac{\mu}{\sigma} E (\min(a, W^* - W) 1_{\{z \leq W \leq z + a\}}) \geq \frac{\mu}{\sigma} E (W^* - W) 1_{\{W^* - W \leq a\}} 1_{\{z \leq W \leq z + a\}}.$$

**Theorem 2.1** Let $Y$ be a nonnegative random variable with mean $\mu$ and variance $\sigma^2$, both finite and positive, and let $Y^*$ be given on the same space as $Y$, with the $Y$ size biased distribution, satisfying $Y \leq Y^* \leq Y + B$ for some $B > 0$. Then with $W$ and $W^*$ given in (12), we have

$$\sup_{z \in \mathbb{R}} |P(W \leq z) - P(Z \leq z)| \leq \frac{\mu}{\sigma^2} \Delta + 0.82 \frac{\delta^2 \mu}{\sigma} + \delta,$$

where

$$\Delta = \sqrt{\operatorname{Var}(E(Y^* - Y|Y))} \quad \text{and} \quad \delta = B/\sigma. \quad (13)$$
Proof. For $z \in \mathbb{R}$ arbitrary, let $h(w) = 1_{\{w \leq z\}}$ and let $f(w)$ be the unique bounded solution to the Stein equation

$$f'(w) - wf(w) = h(w) - Nh,$$

where $Nh = Eh(Z)$ with $Z \sim \mathcal{N}(0, 1)$. Substituting $W$ into (14) and using (12) and (10) yields

$$E(h(W) - Nh) = E(f'(W) - Wf(W))$$

and, as noted in [7], as a consequence of (16) and the mean value theorem, we obtain

$$Nh = Eh(W) = \mu$$

where

$$Nh = Ew = \sigma E$$

and therefore, taking expectation by conditioning, and then applying (16), we may bound the first term in (15) as

$$\int_0^{W^s - W} (f'(W + t) - f'(W)) dt$$

and, as noted in [7], as a consequence of (16) and the mean value theorem, we obtain

$$|(w + t)f(w + t) - w f(w)| \leq (|w| + \sqrt{2\pi}/4)|t|.$$

Noting that $EY^s = EY^2/\mu$ by (10), we find

$$\frac{\mu}{\sigma} E(W^s - W) = \frac{\mu}{\sigma^2} \left( \frac{EY^2}{\mu} - \mu \right) = 1,$$

and therefore, taking expectation by conditioning, and then applying (16), we may bound the first term in (15) as

$$\left| E \left\{ f'(W) E \left( 1 - \frac{\mu}{\sigma} (W^s - W) \right) |W| \right\} \right| \leq \frac{\mu}{\sigma} \sqrt{\text{Var}(E(W^s - W)|W)) = \frac{\mu}{\sigma^2} \Delta.$$ 

To bound the remaining term of (15), using (14), we have

$$\int_0^{W^s - W} (f'(W + t) - f'(W)) dt$$

and, as noted in [7], as a consequence of (16) and the mean value theorem, we obtain

$$|(w + t)f(w + t) - w f(w)| \leq (|w| + \sqrt{2\pi}/4)|t|.$$

Applying (17) to the first term in (18), and using $0 \leq W^s - W \leq \delta$ and $EW^2 = 1$, shows that the absolute value of the expectation of this term is bounded by

$$\frac{\mu}{\sigma} E \left( \int_0^{W^s - W} \left( |W| + \frac{\sqrt{2\pi}}{4} \right) dt \right) \leq \frac{\mu}{2\sigma} E \left( (W^s - W)^2 (|W| + \frac{\sqrt{2\pi}}{4}) \right) \leq \frac{\mu}{2\sigma} \delta^2 (1 + \frac{\sqrt{2\pi}}{4}) \leq 0.82 \frac{\delta^2 \mu}{\sigma}.$$ 

Taking the expectation of the absolute value of the second term in (18), we obtain

$$\left| \frac{\mu}{\sigma} \int_0^{W^s - W} (1_{\{W + t \leq z\}} - 1_{\{W \leq z\}}) dt \right| = \frac{\mu}{\sigma} E \left( \int_0^{W^s - W} 1_{\{z - t < W \leq z\}} dt \right)$$

since $0 \leq W^s - W \leq \delta$ with probability 1. Lemma 2.1 with $a = \delta$ and $z$ replaced by $z - \delta$ shows this term can be no more than $\delta$. Since $z \in \mathbb{R}$ was arbitrary the proof is complete. \qed
3 Normal approximation of \( X \): even case

In this section we provide the proof of Theorem 1.1 when \( n \) is even, starting with a coupling of \( X \), the total number of bulbs on at the terminal time \( n \), to a variable \( X^* \) with the \( X \) size bias distribution. Let \( \mathcal{U}(S) \) denote the uniform distribution over a finite set \( S \).

**Theorem 3.1** With \( n \in \mathbb{N} \) even, let \( X = \{X_{rk} : r, k = 1, \ldots, n\} \) be a collection of switch variables with distribution given by (7) with \( s_r = r \) for \( r = 1, \ldots, n \). For every \( i = 1, \ldots, n \) let \( X^i \) be given from \( X \) as follows. If \( X_i = 1 \) then \( X^i = X \). Otherwise, with \( J^i \sim \mathcal{U}\{j : X_{n/2,j} = 1 - X_{n/2,i}\} \), independent of \( \{X_{rk} : r \neq n/2, k = 1, \ldots, n\} \), let \( X^i = \{X^i_{rk} : r, k = 1, \ldots, n\} \) where

\[
X^i_{rk} = \begin{cases} 
X_{rk} & r \neq n/2 \\
X_{n/2,k} & r = n/2, k \notin \{i, J^i\} \\
X_{n/2,j^i} & r = n/2, k = i \\
X_{n/2,i} & r = n/2, k = J^i 
\end{cases}
\]

and let \( X^i = \sum_{k=1}^n X^i_k \) where

\[
X^i_k = \left( \sum_{r=1}^n X^i_{rk} \right) \mod 2.
\]

Then, with \( I \) uniformly chosen from \( \{1, \ldots, n\} \) and independent of all other variables, the mixture \( X^I = X^* \) has the \( X \) size biased distribution and satisfies

\[
X^* - X = 21_{\{X_i=0, X_{J^i}=0\}}.
\]

In particular, \( X \leq X^* \leq X + 2 \).

We prove Theorem 3.1 making use of a number of preliminary lemmas, and also of the following simple observation. As the marginal distribution of the switch variables is given by

\[
P(X_{ri} = 1) = \frac{r}{n}, \quad \text{for all } r, i = 1, \ldots, n,
\]

when \( n \) is even we have \( P(X_{n/2,i} = 1) = 1/2 \). Hence, by the independence of the switch variables over different stages, for any \( e_1, \ldots, e_n \in \{0, 1\} \),

\[
P(X_{1i} = e_1, \ldots, X_{n/2,i} = e_{n/2}, \ldots, X_{ni} = e_n) = P(X_{1i} = e_1, \ldots, X_{n/2,i} = 1 - e_{n/2}, \ldots, X_{ni} = e_n).
\]

Writing \( x \equiv y \) when \( x = y \mod 2 \), we have in particular

\[
P(X_i = 0) = \sum_{\sum_r e_r = 0} P(X_{1i} = e_1, \ldots, X_{n/2,i} = e_{n/2}, \ldots, X_{ni} = e_n)
\]

\[
= \sum_{\sum_r e_r = 0} P(X_{1i} = e_1, \ldots, X_{n/2,i} = 1 - e_{n/2}, \ldots, X_{ni} = e_n)
\]

\[
= \sum_{\sum_r e_r = 1} P(X_{1i} = e_1, \ldots, X_{n/2,i} = e_{n/2}, \ldots, X_{ni} = e_n)
\]

\[
= P(X_i = 1),
\]

so

\[
P(X_i = 0) = P(X_i = 1) = \frac{1}{2} \quad \text{for all } i = 1, \ldots, n.
\]

**Lemma 3.1** For all \( i = 1, \ldots, n \), the collections of random variables \( X \) and \( X^i \) as specified in Theorem 3.1 satisfies

\[
\mathcal{L}(X^i) = \mathcal{L}(X|X_i = 1).
\]
Proof. Let \( i \in \{1, \ldots, n\} \) and \( e^i = \{ e^i_{rk} : r, k = 1, \ldots, n \} \) with \( e^i_{rk} \in \{0, 1\} \) for \( r, k = 1, \ldots, n \). First note that

\[
P(X^i = e^i) = \frac{1}{2} P(X^i = e^i_i | X^i_i = 1) + \frac{1}{2} P(X^i = e^i_i | X^i_i = 0)
\]

so it suffices to prove

\[
P(X^i = e^i_i | X^i_i = 0) = P(X = e^i_i | X^i_i = 1).
\] (21)

As the construction of \( X^i \) preserves the number of switches in each stage, that is, since \( \sum_k X^i_{rk} = \sum_k e^i_{rk} = r \) for all \( r \), as otherwise both sides of (21) are zero. By construction we have \( X^i_1 = 1 \), so \( P(X^i = e^i_i) = 0 \) whenever \( \sum_r e^i_{ri} \equiv 0 \), in which case, again, both sides of (21) are zero. Hence we need only verify (21) assuming

\[
\sum_{r=1}^n e^i_{ri} \equiv 1. 
\] (22)

To look at the values of the switch variables for bulb \( k, k = 1, \ldots, n \) over all stages, let \( e^i_k = (e^i_{1k}, \ldots, e^i_{nk}), \ X_k = (X^i_1, \ldots, X^i_n) \) and \( X^i_k = (X^i_{1k}, \ldots, X^i_{nk}) \) for \( k = 1, \ldots, n \).

When (22) holds, writing \( J \) for \( J^i \) for simplicity, we have

\[
P(X^i = e^i_i | X^i_i = 0) = 2 \sum_{j=1}^n P(X^i_k = e^i_k, k \notin \{i, j\}, X^i_i = e^i_j, X^i_j = e^i_j, X^i_i = 0, J = j).
\]

On the event \( J = j \) the vectors \( X^i_k \) and \( X^i_j \) are equal to \( X^i_i \) and \( X^i_j \), respectively, with their \( n/2^{nd} \) coordinates interchanged, and the values at these coordinates are unequal. Hence, letting \( e^i_{ij} \) and \( e^i_{ji} \) equal \( e^i_i \) and \( e^i_j \), respectively, with their \( n/2^{nd} \) coordinates interchanged, we have

\[
P(X^i = e^i_i | X^i_i = 0) = 2 \sum_{j=1}^n P(X^i_k = e^i_k, k \notin \{i, j\}, X^i_i = e^i_{ij}, X^i_j = e^i_{ji}, X^i_i = 0, J = j).
\]

Note that the probability above is zero for \( e^i_{ij}/e^i_{ji} = e^i_{n/2,i} \) and when \( J_i = j \) the statuses of \( X^n_{n/2,i} \) and \( X^n_{n/2,j} \) are different, and when \( e^i_{n/2,i} \neq e^i_{n/2,j} \) and \( X^i_i = e^i_{ij} \), then by (22) we have

\[
X^i_i \equiv \sum_{r=1}^n e^i_{ri} \equiv \sum_{r=1}^n e^i_{ri} + 1 \equiv 0.
\]

Hence,

\[
P(X^i = e^i_i | X^i_i = 0) = 2 \sum_{j=1}^n P(X^i_k = e^i_k, k \notin \{i, j\}, X^i_i = e^i_{ij}, X^i_j = e^i_{ji}, J = j)
\]

\[
= 2 \sum_{j=1}^n P(J = j) P(X^i_k = e^i_k, k \notin \{i, j\}, X^i_i = e^i_{ij}, X^i_j = e^i_{ji}| J = j).
\]

8
Since $J$ is independent of the variables in all stages $r \neq n/2$, the second probability factors, and the $j^{th}$ summand equals
\[
\prod_{r \neq n/2} P(X_{rk} = e_{rk}^i, k \notin \{i, j\}, X_{ri} = e_{ri}^j, X_{rj} = e_{rj}^j) \\
\times P(J = j) P(X_{n/2,k} = e_{n/2,k}^i, k \notin \{i, j\}, X_{n/2,i} = e_{n/2,i}^j, X_{n/2,j} = e_{n/2,j}^j | J = j)
\]
\[
= \prod_{r \neq n/2} P(X_{rk} = e_{rk}^i, k = 1, \ldots, n) \\
\times P(J = j) P(X_{n/2,k} = e_{n/2,k}^i, k \notin \{i, j\}, X_{n/2,i} = e_{n/2,i}^j, X_{n/2,j} = e_{n/2,j}^j | J = j),
\]
the last equality due to the fact that $e_{rk}^i$ and $e_{rk}^j$ equal $e_{rk}^i$ for $r \neq n/2$ and $k \notin \{i, j\}$, and that $e_{n/2,i}^j = e_{n/2,j}^i$ and $e_{n/2,j}^j = e_{n/2,i}^i$. As the first product does not depend on $j$, taking it past the sum we obtain
\[
P(X^i = e^i | X_i = 0) = \prod_{r \neq n/2} P(X_{rk} = e_{rk}^i, k = 1, \ldots, n)
\]
\[
\times 2 \sum_{j=1}^n P(X_{n/2,k} = e_{n/2,k}^i, k \notin \{i, j\}, X_{n/2,i} = e_{n/2,i}^j, X_{n/2,j} = e_{n/2,j}^j, J = j).
\]

Since for each $j$ the $j^{th}$ summand in (23) equals zero when $e_{n/2,i}^j = e_{n/2,j}^i$, and also noting that there are $n/2$ indices whose switch variable at state $n/2$ is of opposite status to that of the $i^{th}$, the $j^{th}$ summand equals
\[
P(X_{n/2,k} = e_{n/2,k}^i, k \notin \{i, j\}, X_{n/2,i} = e_{n/2,i}^j, X_{n/2,j} = e_{n/2,j}^j, J = j) 1(e_{n/2,i}^j \neq e_{n/2,j}^i)
\]
\[
= P(J = j) P(X_{n/2,k} = e_{n/2,k}^i, k \notin \{i, j\}, X_{n/2,i} = e_{n/2,i}^j, X_{n/2,j} = e_{n/2,j}^j) 1(e_{n/2,i}^j \neq e_{n/2,j}^i)
\]
\[
\times 2 \sum_{j=1}^n P(X_{n/2,k} = e_{n/2,k}^i, k \notin \{i, j\}, X_{n/2,i} = e_{n/2,i}^j, X_{n/2,j} = e_{n/2,j}^j, J = j).
\]

These probabilities are equal for all the $n/2$ values of $j$ for which $e_{n/2,i}^j \neq e_{n/2,j}^i$, and zero otherwise, so summing over $j$ in (23) yields, now taking $j$ to be any index such that $e_{n/2,i}^j \neq e_{n/2,j}^i$:
\[
P(X^i = e^i | X_i = 0) = \prod_{r \neq n/2} P(X_{rk} = e_{rk}^i, k = 1, \ldots, n)
\]
\[
\times 2^2 P(X_{n/2,k} = e_{n/2,k}^i, k \notin \{i, j\}, X_{n/2,i} = e_{n/2,i}^j, X_{n/2,j} = e_{n/2,j}^j)
\]
\[
= \prod_{r \neq n/2} P(X_{rk} = e_{rk}^i, k = 1, \ldots, n)
\]
\[
\times 2^n P(X_{n/2,k} = e_{n/2,k}^i, k \notin \{i, j\}, X_{n/2,i} = e_{n/2,i}^j, X_{n/2,j} = e_{n/2,j}^j)
\]
\[
= 2 \prod_{r=1}^n P(X_{rk} = e_{rk}^i, k = 1, \ldots, n) = 2 P(X = e^i),
\]
where we have used exchangeability to obtain the second equality.

But now, by (22),
\[
P(X^i = e^i | X_i = 0) = 2 P(X = e^i) = 2 P(X = e^i, X_i = 1) = P(X = e^i | X_i = 1),
\]
which is (21).}

The next lemma will be used to construct our even case coupling. The result is a special case of Lemma 2.1 of [4], but we give a short direct proof to make the paper more self-contained.
Lemma 3.2 Suppose \( X \) is a sum of nontrivial exchangeable Bernoulli variables \( X_1, \ldots, X_n \), and that for \( i \in \{1, \ldots, n\} \) the variables \( X^i \) have joint distribution
\[
\mathcal{L}(X^i) = \mathcal{L}(X_1, \ldots, X_n|X_i = 1).
\]

Then
\[
X^i = \sum_{j=1}^{n} X^i_j,
\]
has the size biased distribution \( X^* \), as does the mixture \( X^j \) when \( I \) is a random index with values in \( \{1, \ldots, n\} \), independent of all other variables.

Proof. We need to show that \( X^i \) satisfies (10), that is, that \( E[X|Eg(X^i)] = E[Xg(X)] \) holds for a given bounded continuous \( g \). Now, for such \( g \)
\[
E[Xg(X)] = \sum_{j=1}^{n} E[X_jg(X)] = \sum_{j=1}^{n} P[X_j = 1]E[g(X)|X_j = 1].
\]
As exchangeability implies that \( E[g(X)|X_j = 1] \) does not depend on \( j \), we have
\[
E[Xg(X)] = \left( \sum_{j=1}^{n} P[X_j = 1] \right) E[g(X)|X_i = 1] = E[X]E[g(X^i)],
\]
demonstrating the first result. The second follows from
\[
Ef(X^I) = \sum_{i=1}^{n} Ef(f(X^I), I = i) = \sum_{i=1}^{n} Ef(f(X^I)|I = i)P(I = i) = \sum_{i=1}^{n} Ef(X^i)P(I = i)
\]
\[
= \sum_{i=1}^{n} Ef(X^*)P(I = i) = Ef(X^*)\sum_{i=1}^{n} P(I = i) = Ef(X^*).\]

We now present the proof of Theorem 3.1.

Proof. With \( X^i \) as given in the theorem, Lemma 3.1 yields that the hypotheses of Lemma 3.2 are satisfied, so we may conclude that the given \( X^* \) has the size biased distribution. To prove (19), first note that if \( X_I = 1 \) then \( X^I = X \), hence in this case \( X^* = X \). Otherwise, for \( X_I = 0 \), recall that for the given \( I \) the collection \( X^I \) is constructed from \( X \) by interchanging the stage \( n/2 \), unequal, switch variables \( X^i_{n/2,J} \) and \( X^j_{n/2,I} \). If \( X^i_{J} = 1 \) then after the interchange \( X^i_{J} = 1 \) and \( X^j_{J} = 0 \), yielding \( X^* = X \). If \( X^j_{I} = 0 \) then after the interchange \( X^j_{I} = 1 \) and \( X^i_{I} = 1 \), yielding \( X^* = X + 2 \).

Based on the spectral decomposition in the Appendix, we now provide an upper bound to the term (13) in Theorem 2.1, for this case.

Lemma 3.3 Let \( n \) be even and \( X \) and \( X^* \) given by Theorem 3.1. Then for \( n \geq 6 \)
\[
\Delta_0 \leq \frac{1}{2\sqrt{n}} + \frac{1}{2n} + e^{-n/2} \quad \text{where} \quad \Delta_0 = \sqrt{\text{Var}(E(X^* - X|X))}.
\]

Proof. Recall the construction of Theorem 3.1, and for notational simplicity let \( I^j = J \), so that we may write \( X^* - X = 21_{\{X_I = 0, X_J = 0\}} \). Expanding the indicator over the possible values of \( I \) and \( J \), and then over the values of the relevant switch variables at stage \( n/2 \), we have
\[
1_{\{X_I = 0, X_J = 0\}} = \sum_{i,j=1}^{n} 1_{\{X_i = 0, X_j = 0\}} 1_{\{I = i, J = j\}}
\]
\[ = \sum_{i,j=1}^{n} 1\{X_i=0,X_j=0,X_{n/2,i}=0\} 1\{I=i,J=j\} + \sum_{i,j=1}^{n} 1\{X_i=0,X_j=0,X_{n/2,i}=1\} 1\{I=i,J=j\} \]
\[ = \sum_{i \neq j} 1\{X_i=0,X_j=0,X_{n/2,i}=0\} 1\{I=i,J=j\} + \sum_{i \neq j} 1\{X_i=0,X_j=0,X_{n/2,i}=1\} 1\{I=i,J=j\} \]
\[ = 2 \sum_{i \neq j} 1\{X_i=0,X_j=0,X_{n/2,i}=0\} 1\{I=i,J=j\}, \]

where the second to last equality holds almost surely, as the probability of the event \(\{I = i, J = j\}\) is zero whenever \(X_{n/2,i}\) and \(X_{n/2,j}\) agree, and the last inequality holds as the final expression is the sum of two terms which remain the same when interchanging \(i\) and \(j\).

Letting \(\mathcal{F}\) be the \(\sigma\)-algebra generated by \(X\), the collection of all switch variables. The first term in the sum above, \(1\{X_i=0,X_j=0,X_{n/2,i}=0\} 1\{I=i,J=j\}\), is measurable with respect to \(\mathcal{F}\), while for the second conditioning yields
\[ P(I = i, J = j|\mathcal{F}) = \frac{2}{n^2} 1\{X_{n/2,i} \neq X_{n/2,j}\}, \]

as for any \(i\), there are \(n/2\) choices for \(j\) satisfying the condition in the indicator above. Hence, recalling that \(X^s - X = 21\{X_j=0,X_j=0\}\), we have
\[ E\left(X^s - X \mid \mathcal{F}\right) = U_n \text{ where } U_n = \frac{4}{n^2} \sum_{i \neq j} 1\{X_i=0,X_j=0,X_{n/2,i}=0,X_{n/2,j}=1\}, \quad (24) \]

and
\[ \Delta_0^2 = \text{Var}(U_n). \]

Taking the expectation of \(U_n\) in (24), by the exchangeability of the \((n)\) terms in the sum and identity (58),
\[ EU_n = \frac{4}{n^2} (n) g_{1,1,n|[n]/n/2} = \frac{1}{4} \left(1 - \lambda_{n,2|[n]/n/2}\right). \quad (25) \]

Squaring (24) in order to obtain the second moment of \(U_n\), we obtain a sum over indices \(i_1, i_2, j_1, j_2\) with \(i_1 \neq j_1, i_2 \neq j_2\), and \(i_1 \neq j_2, i_2 \neq j_1\), so \(|\{i_1, i_2, j_1, j_2\}| \in \{2, 3, 4\}\), and we may write
\[ U_n^2 = U_{n,2}^2 + U_{n,3}^2 + U_{n,4}^2 \quad \text{where} \]
\[ U_{n,p}^2 = \frac{16}{n^2} \sum_{\{i_1, i_2, j_1, j_2\} \neq p} 1\{X_{i_1}=0,X_{j_1}=0,X_{n/2,i_1}=0,X_{n/2,j_1}=1\} 1\{X_{i_2}=0,X_{j_2}=0,X_{n/2,i_2}=0,X_{n/2,j_2}=1\}. \quad (26) \]

Beginning the calculation with the main term \(U_{n,4}^2\), where all four indices are distinct, taking expectation and using exchangeability and (59) yields
\[ EU_{n,4}^2 = \frac{16}{n^2} (n) g_{2,2,n|[n]/n/2} = \left(\frac{n-2}{4n}\right)^2 \left(1 - 2\lambda_{n,2|[n]/n/2} + \lambda_{n,4|[n]/n/2}\right). \quad (27) \]

With the inequalities over the summation in (26) in force, the event \(|\{i_1, i_2, j_1, j_2\}| = 3\) can occur when
\( a) i_1 \neq i_2, j_1 = j_2 \quad \text{or} \quad b) i_1 = i_2, j_1 \neq j_2. \)

Using (60), case \(a)\) leads to a contribution of
\[ \frac{16}{n^2} (n) s g_{2,1,n|[n]/n/2} = \frac{n-2}{4n^2} \left(1 + \lambda_{n,1|[n]/n/2} - \lambda_{n,2|[n]/n/2} - \lambda_{n,3|[n]/n/2}\right), \]

11
while in the same manner, using (61), the contribution from case b) is
\[
\frac{16}{n^4} g_{1,2,n,n/2,n/2} = \left(1 - \lambda_{n,1,[n]/n} - \lambda_{n,2,[n]/n} + \lambda_{n,3,[n]/n}\right).
\]

Totaling we find
\[
EU_{n,2}^2 = \frac{n - 2}{2n^2} \left(1 - \lambda_{n,2,[n]/n}\right).
\]  \hspace{1cm} (28)

With the inequalities over the summation in (26) in force, the event \(|\{i_1,i_2,j_1,j_2\}| = 2\) can only occur when \(i_1 = i_2\) and \(j_1 = j_2\), hence, again using (58),
\[
EU_{n,2}^2 = \frac{16}{n^4} g_{1,1,1,n,n/2,n/2} = \frac{1}{n^2} \left(1 - \lambda_{n,2,[n]/n}\right).
\]  \hspace{1cm} (29)

Summing (27), (28) and (29),
\[
EU_n^2 = \left(\frac{n - 2}{4n}\right)^2 \left(1 - 2\lambda_{n,2,[n]/n} + \lambda_{n,4,[n]/n}\right) + \frac{1}{2n} \left(1 - \lambda_{n,2,[n]/n}\right).
\]

Hence, by (25),
\[
\begin{align*}
\text{Var}(U_n) &= \frac{1}{16} \left(1 - \frac{2}{n}\right)^2 \left(1 - 2\lambda_{n,2,[n]/n} + \lambda_{n,4,[n]/n}\right) + \frac{1}{2n} \left(1 - \lambda_{n,2,[n]/n}\right) - \frac{1}{16} \left(1 - \lambda_{n,2,[n]/n}\right)^2 \\
&= \frac{1}{16} \left(\lambda_{n,4,[n]/n} - \lambda_{n,2,[n]/n}\right)^2 + \frac{1 - n}{4n^2} \left(1 - 2\lambda_{n,2,[n]/n} + \lambda_{n,4,[n]/n}\right) + \frac{1}{2n} \left(1 - \lambda_{n,2,[n]/n}\right) \\
&= \frac{1}{4n} + \frac{1}{4n^2} + \frac{1}{16} \left(\lambda_{n,4,[n]/n} - \lambda_{n,2,[n]/n}\right)^2 - \frac{1}{2n^2} \lambda_{n,2,[n]/n} + \frac{1 - n}{4n^2} \lambda_{n,4,[n]/n}.
\end{align*}
\]

Now applying Lemma 5.3, for \(n \geq 6\),
\[
\text{Var}(U_n) \leq \frac{1}{4n} + \frac{1}{4n^2} + \frac{1}{16} \left(\frac{1}{2} e^{-n} + e^{-2n}\right) + \frac{1}{2n^2} e^{-n} + \frac{1 - n}{8n^2} e^{-n}
\]
\[
\leq \frac{1}{4n} + \frac{1}{4n^2} + e^{-n} \left(\frac{1}{16} + \frac{1}{n^2} + \frac{1}{8n}\right),
\]

and applying the inequality \(\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}\) for nonnegative \(a\) and \(b\) yields the claim of the lemma. \(\square\)

With all ingredients at hand, we may now prove the bound for even \(n\).

Proof of Theorem 1.1. even case. The size biased coupling given in Theorem 3.1 satisfies the hypotheses of Theorem 2.1 with \(B = 2\), and the result for the even case follows by applying Theorem 2.1 with \(\delta = 2/\sigma\) and the bound in Lemma 3.3. \(\square\)

4 Normal Approximation of \(X\): Odd case

Now we move to the case where \(n = 2m + 1\) is odd. Instead of directly forming a size biased coupling of \(X^\ast\) to \(X\), we first couple \(X\) closely to a more symmetrical random variable \(V\), for which a coupling like the one in the even case may be applied. The variable \(V\) is constructed by randomizing stages \(m\) and \(m + 1\) as follows. With equal probability we either add an additional switch at stage \(m\) or remove a switch at stage \(m + 1\). We prove a normal bound for \(V\) in the same way as for \(X\) in the even case, and may then derive the normal bound for \(X\) in the odd case based on \(X\)’s proximity to \(V\).

Formally, let \(X\) be given with distribution (7) with \(s_r = r, r = 1, \ldots, n\) and set
\[
J_r = \{j : X_{r,j} = 0\} \text{ for } r = m, m + 1,
\]
so that $\mathcal{J}_m$ and $\mathcal{J}_{m+1}$ are the bulbs that do not get toggled in stages $m$ and $m+1$, respectively. Let $N \sim \mathcal{U}(m, m+1)$, $B_m \sim \mathcal{U}(\mathcal{J}_m)$ and $B_{m+1} \sim \mathcal{U}(\mathcal{J}_{m+1})$, independent of $X$ and of each other. Now let the collection $V$ be given by

$$
V_{rk} = \begin{cases} 
X_{rk} & r \neq N \\
X_{mk} & N = m, k \neq B_m \\
1 & N = m, k = B_m \\
X_{m+1,k} & N = m+1, k \neq B_{m+1} \\
0 & N = m+1, k = B_{m+1},
\end{cases}
$$

\hspace{1cm} (30)

and

$$
V = \sum_{k=1}^{n} V_k \text{ where } V_k = \left( \sum_{r=1}^{n} V_{rk} \right) \text{ mod } 2.
$$

\hspace{1cm} (31)

In other words, in all stages other than stage $N$ the switch variables that produce $V$ are the ones from the given collection $X$. If $N = m$, then the switch variables for all bulbs but bulb $B_m$, chosen uniformly over all bulbs in that stage that were not toggled, are the ones given by $X$ and the switch variable for $B_m$ is set to 1. Similarly, if $N = m+1$, the switch variable of bulb $B_{m+1}$, uniformly selected from all the bulbs that were toggled in that stage, is no longer toggled.

Various other couplings are possible, with mixed effects. In particular, the same analysis as the one below can be applied to the scheme where in stage $m$ with probability 1/2 one additional switch is used and, independently in stage $m + 1$, with probability 1/2, one fewer. As two variables are affected the bound of 1 in (34) increases to 2, but, on the other hand, more symmetry results, and, in particular, the variable so constructed has mean zero. See [2] where the latter coupling was used to obtain a moderate deviation bound for $X$.

Recalling the definitions of $s_{m,m}$ and $s_{m+1,m+1}$ in (3), we may express the distribution of the collection of switch variables $V$ succinctly as

$$
V =_d \begin{cases} 
X_{s_{m,m}} & N = m+1 \\
X_{s_{m+1,m+1}} & N = m,
\end{cases}
$$

\hspace{1cm} (32)

where $X_{s_{m,m}}$ and $X_{s_{m+1,m+1}}$ are switch variables with distribution (7), and $=_d$ denotes equality in distribution. As $N \sim \mathcal{U}(m, m+1)$, the law $\mathcal{L}(V)$ of $V$ is the equal mixture of

$$
\mathcal{L}(V|N = m+1) = \mathcal{L}(X_{s_{m,m}}) \text{ and } \mathcal{L}(V|N = m) = \mathcal{L}(X_{s_{m+1,m+1}}).
$$

\hspace{1cm} (33)

Consequently, for $r \in \{m, m+1\}$ and $\{r, -r\} = \{m, m+1\}$,

$$
P(V_{r1} = e_1, \ldots, V_{rn} = e_n|N = r) = P(X_{-r,1} = e_1, \ldots, X_{-r,n} = e_n)
$$

$$
P(V_{r1} = e_1, \ldots, V_{rn} = e_n|N = -r) = P(X_{r1} = e_1, \ldots, X_{rn} = e_n),
$$

and

$$
\mathcal{L}(V_{1}, \ldots, V_{n}) = \frac{1}{2} \mathcal{L}(X_{m1}, \ldots, X_{mn}) + \frac{1}{2} \mathcal{L}(X_{m+1,1}, \ldots, X_{m+1,n}).
$$

Arguing as for (20),

$$
P(V_{r1} = 1) = \frac{1}{2} \text{ for } r \in \{m, m+1\} \text{ which implies } P(V_{i} = 1) = \frac{1}{2} \text{ for all } i = 1, \ldots, n.
$$

The following theorem shows how to construct a monotone, bounded size bias coupling to $V$. 

13
Theorem 4.1 With \( n = 2m + 1 \) let \( X \) have distribution (7) with \( s_r = r, r = 1, \ldots, n \) and let \( V = \{ V_{rk} : r, k = 1, \ldots, n \} \) be constructed from \( X \) as in (30). For every \( i = 1, \ldots, n \) let \( V^i \) be given from \( V \) as follows. If \( V_i = 1 \) then let \( V^i = V \). Otherwise, with \( M \sim U\{m, m + 1\} \) and \( J^i = U\{j : V_{M,j} = 1 - V_{M,i}\} \) let \( V^i = \{ V^i_{rk} : r, k = 1, \ldots, n \} \) where

\[
V^i_{rk} = \begin{cases} 
V_{rk} & r \neq M \\
V_{M,k} & r = M, k \notin \{i, J^i\} \\
V_{M,i} & r = M, k = i \\
V_{M,i} & r = M, k = J^i,
\end{cases}
\]

and let \( V^i = \sum_{k=1}^n V^i_k \) where

\[
V^i_k = \left( \sum_{j=1}^r V^i_{jk} \right) \mod 2.
\]

Then, with \( I \) uniformly chosen from \( \{1, \ldots, n\} \) and independent of all other variables, the mixture \( V^s = V^I \) has the \( V \) size biased distribution and satisfies

\[
V^s - V = 21_{\{V_i = 0, V_{J,i} = 0\}},
\]

so in particular \( V \leq V^s \leq V + 2 \). In addition, with probability one,

\[
|X - V| \leq 1,
\]

where \( X \), given by (8) with \( s = (1, \ldots, n) \), is the number of bulbs on at the terminal time using switch variables \( X \).

In other words, in stage \( m \) or stage \( m + 1 \), as determined by \( M \), when \( V_i = 0 \) the switch for bulb \( i \) at stage \( M \) is interchanged with one having opposite parity.

The proof of Theorem 4.1 follows from the Lemma 4.1, as in the even case.

Lemma 4.1 With \( V^i \) constructed as in Theorem 4.1,

\[
\mathcal{L}(V^i) = \mathcal{L}(V|V_i = 1).
\]

Proof. We argue as in Lemma 3.1, highlighting the differences, and use parallel notation, such as writing \( J \) for \( J^i \). It again suffices to show

\[
P(V^i = e^i|V_i = 0) = P(V = e^i|V_i = 1)
\]

for

\[
\sum_{k=1}^n e^i_{rk} = \begin{cases} 
r & r \notin \{m, m + 1\} \\
m \text{ or } m + 1 & r \in \{m, m + 1\},
\end{cases}
\]

and

\[
\sum_{r=1}^n e^i_{ri} = 1.
\]

Forming the vector of switch variables for bulb \( k \),

\[
V_k = (V_{1k}, \ldots, V_{nk})
\]

and defining vectors such as \( e^i_k \) likewise, using that \( P(V_i = 0) = 1/2 \) and that \( V^i_k = V_k \) for \( k \notin \{i, J\} \), decomposing by the values taken on by \( J \) and \( M \), we have

\[
P(V^i = e^i|V_i = 0) = 2 \sum_{r=m}^{m+1} \sum_{j=1}^n P(V_k = e^i_k, k \notin \{i, j\}, V^i_j = e^i_j, V_i = 0, J = j, M = r).
\]
On the event \( \{ J = j, M = r \} \) the vectors \( V_i \) and \( V_j \) are equal to \( V_i \) and \( V_j \) with their \( r \)th coordinates interchanged, and the values at these coordinates are unequal. Hence, letting \( e_{ijr} \) and \( e_{jir} \) equal \( e_i \) and \( e_j \), respectively, with their \( r \)th coordinates interchanged, the above probability is

\[
2 \sum_{r=m}^{m+1} \sum_{j=1}^{n} P(V_k = e_{ik}, k \notin \{i, j\}, V_i = e_{ijr}, V_j = e_{jir}, V_i = 0, J = j, M = r).
\]

Since the probabilities above are zero if \( e_{ri} = e_{rj} \), and that \( V_i^j = 1 \) implies \( V_i = 0 \) when \( e_{ri} \neq e_{rj} \), the sum equals

\[
2 \sum_{r=m}^{m+1} \sum_{j=1}^{n} P(V_k = e_{ik}, k \notin \{i, j\}, V_i = e_{ijr}, V_j = e_{jir}, J = j, M = r)
= 2 \sum_{r=m}^{m+1} \sum_{j=1}^{n} P(J = j, M = r)P(V_k = e_{ik}, k \notin \{i, j\}, V_i = e_{ijr}, V_j = e_{jir} | J = j, M = r).
\]

By the independence of the switch variables in stages \( s \notin r \) of \( \{ J = j, M = r \} \), the \( r, j \)th summand may be written

\[
\prod_{s \notin r} P(V_{sk} = e_{sk}, k \notin \{i, j\}, V_{si} = e_{sir}, V_{sj} = e_{sjr})
\times P(J = j, M = r)P(V_k = e_{rk}, k \notin \{i, j\}, V_{ri} = e_{rjr}, V_{rj} = e_{rjr} | J = j, M = r)
= \prod_{s \notin r} P(V_{sk} = e_{sk}, k = 1, \ldots, n)
\times P(V_{rk} = e_{rk}, k \notin \{i, j\}, V_{ri} = e_{rjr}, V_{rj} = e_{rjr}, J = j, M = r),
\]

the last equality due to the fact that \( e_{sir} \) and \( e_{sjr} \) equal \( e_{sk} \) for \( s \neq r \), and otherwise are \( e_{rjr} \) and \( e_{rjr} \), respectively. As the first product does not depend on \( j \), we may write

\[
P(V_i = e_i | V_i = 0) = 2 \sum_{r=m}^{m+1} \left( \prod_{s \notin r} P(V_{sk} = e_{sk}, k = 1, \ldots, n) \right.
\times \sum_{j=1}^{n} P(V_{rk} = e_{rk}, k \notin \{i, j\}, V_{ri} = e_{rjr}, V_{rj} = e_{rjr}, J = j, M = r)) \right)
\]

Further decomposing according to the value of \( N \), the term in the final sum equals

\[
P(V_{rk} = e_{rk}, k \notin \{i, j\}, V_{ri} = e_{rjr}, V_{rj} = e_{rjr}, J = j, M = r)
= P(V_{rk} = e_{rk}, k \notin \{i, j\}, V_{ri} = e_{rjr}, V_{rj} = e_{rjr}, J = j, M = r, N \neq r)
+ P(V_{rk} = e_{rk}, k \notin \{i, j\}, V_{ri} = e_{rjr}, V_{rj} = e_{rjr}, J = j, M = r, N = r).
\]

The two terms above must be handled by considering further cases, depending on the values of \( r, e_{ri} \) and \( e_{rj} \). Note first that the two probabilities being summed are zero when \( e_{ri} = e_{rj} \), so we consider only \( e_{ri} \neq e_{rj} \). For instance, for the first term in (37), specializing to \( r = m \) and \( e_{mi} = 0, e_{mj} = 1 \) results in

\[
P(V_{mk} = e_{mk}, k \notin \{i, j\}, V_{mi} = 0, V_{mj} = 1, J = j, M = m, N = m + 1)1(e_{mi} = 0, e_{mj} = 1)
= P(J = j | V_{mk} = e_{mk}, k \notin \{i, j\}, V_{mi} = 0, V_{mj} = 1, M = m, N = m + 1)
\times P(V_{mk} = e_{mk}, k \notin \{i, j\}, V_{mi} = 0, V_{mj} = 1, M = m, N = m + 1)1(e_{mi} = 0, e_{mj} = 1).
\]
Note that due to the condition \( N = m + 1 \), the last probability above is 0 unless \( \sum_k e^i_{mk} = m \). When \( \sum_k e^i_{mk} = m \) and \( V_{mi} = 0 \), any \( j \) with status 1 has probability of \( 1/m \) of being chosen as \( J \). Using that \( V = X \) when \( M = m \) and \( N = m + 1 \), and the independence of \( M \) and \( N \) from \( X \), the product above equals

\[
\frac{1}{m} P(V_{mk} = e^i_{mk}, k \notin \{i, j\}, V_{mi} = e^i_{mj}, V_{mj} = e^i_{mi}, M = m, N = m + 1) \mathbf{1}(e^i_{mi} = 0, e^i_{mj} = 1)
\]

\[
= \frac{1}{m} P(X_{mk} = e^i_{mk}, k \notin \{i, j\}, X_{mi} = e^i_{mj}, X_{mj} = e^i_{mi}, M = m, N = m + 1) \mathbf{1}(e^i_{mi} = 0, e^i_{mj} = 1)
\]

\[
= \frac{1}{4m} P(X_{mk} = e^i_{mk}, k \notin \{i, j\}, X_{mi} = e^i_{mj}, X_{mj} = e^i_{mi}) \mathbf{1}(e^i_{mi} = 1, e^i_{mj} = 0).
\]

Summing over \( j \) as indicated in (36) leads to the following contribution from the event \( \{M = m, N = m + 1\} \), taking \( j \) below to be any index not equal to \( i \),

\[
\frac{1}{4} P(X_{mk} = e^i_{mk}, k \notin \{i, j\}, X_{mi} = 0, X_{mj} = 1) \mathbf{1}(e^i_{mi} = 0) \mathbf{1} \left( \sum_k e^i_{mk} = m \right)
\]

\[
+ \frac{1}{4} P(X_{mk} = e^i_{mk}, k \notin \{i, j\}, X_{mi} = 1, X_{mj} = 0) \mathbf{1}(e^i_{mi} = 1) \mathbf{1} \left( \sum_k e^i_{mk} = m \right)
\]

\[
= \frac{1}{4} \left( \frac{n}{m} \right)^{-1} \left( \mathbf{1}(e^i_{mi} = 0) + \mathbf{1}(e^i_{mi} = 1) \right) \mathbf{1} \left( \sum_k e^i_{mk} = m \right)
\]

\[
= \frac{1}{4} P(X_{mk} = e^i_{mk}, k = 1, \ldots, n).
\]

Likewise, the contribution from \( \{M = m, N = m + 1\} \) is \((1/4)P(X_{m+1,k} = e^i_{mk}, k = 1, \ldots, n)\), yielding that the second sum (36), for \( r = m \), equals \((1/2)P(V_{mk} = e^i_{mk}, k = 1, \ldots, n)\). Arguing in like manner for \( r = m + 1 \) to obtain the same expression, substitution into (36) yields

\[
P(V^i = e^i | V_i = 0)
\]

\[
= 2 \sum_{r=m+1}^n P(V_{rk} = e^i_{rk}, k = 1, \ldots, n) P(V_{mk} = e^i_{mk}, k = 1, \ldots, n)
\]

\[
= 2 P(V = e^i | V_i = 1),
\]

by (35).

The last claim (34) of the theorem is immediate upon noting that \( X \) and \( V \) differ by at most one switch variable, the one indexed by \( M, B_M \). Hence \( X_k = V_k \) for all \( k \neq B_M \), so \( X \) and \( V \) can differ by at most 1. \( \square \)

With the coupling now in hand, we prove a bound to the normal for \( V \).

**Theorem 4.2** If \( n = 2m+1 \), an odd number, and \( V \) is given by (31), then

\[
\sup_{z \in \mathbb{R}} \left| P \left( \frac{V - \mu_V}{\sigma_V} \leq z \right) - P(Z \leq z) \right| \leq \frac{\mu_V}{\sigma_V^2} \Delta_1 + 3.28 \frac{\mu_V}{\sigma_V^2} + \frac{2}{\sigma_V},
\]

16
where the mean $EV$ and variance $\text{Var}(V)$ are given by $\mu_V$ and $\sigma^2_V$ in (4), respectively, and $\Delta_1$ satisfies the upper bound in (6).

**Proof.** That $EV = \mu_V$ and $\text{Var}(V) = \sigma^2_V$ follows from (2) and (32), and, for the latter equality, the conditional variance formula.

By Theorem 4.1, we may apply Theorem 2.1 with $\delta = 2/\sigma_V$, and it only remains to prove the upper bound (6) on $\Delta_1 = \sqrt{\text{Var}(V^* - V|V)}$, for which we parallel the calculation of Section 3. Again, for notational simplicity, let $J = J^I$. From Theorem 4.1,

$$V^* - V = 2\mathbf{1}_{\{V_i = 0, V_j = 0\}}.$$

Decomposing based on the possible values of $I, J$,

$$\begin{align*}
1_{\{V_i = 0, V_j = 0\}} &= \sum_{i,j=1}^n 1_{\{V_i = 0, V_j = 0\}} 1_{\{I = i, J = j\}} \\
&= \sum_{i,j=1}^n 1_{\{V_i = 0, V_j = 0, V_{M_i} = 0, I = i, J = j\}} + \sum_{i,j=1}^n 1_{\{V_i = 0, V_j = 0, V_{M_i} = 1, I = i, J = j\}} \\
&= \sum_{i \neq j} 1_{\{V_i = 0, V_j = 0, V_{M_i} = 0, V_{M_j} = 1\}} 1_{\{I = i, J = j\}} + \sum_{i \neq j} 1_{\{V_i = 0, V_j = 0, V_{M_i} = 1, V_{M_j} = 0\}} 1_{\{I = i, J = j\}},
\end{align*}$$

where for the last equality we have used that $J$ always selects a variable with opposite status to $V_{M_i}$.

Letting $F$ be the $\sigma$-algebra generated by $\{V_{rk}, M, N\}$, the first terms in the sums above are measurable with respect to $F$, while for the second, similar to the identity in the even case analysis,

$$P(I = i, J = j|F) = \frac{1}{nm} \left( 1_{\{V_{M_i} = 1, V_{M_j} = 0, N = m\}} + 1_{\{V_{M_i} = 0, V_{M_j} = 1, N = m+1\}} \right)$$

$$+ \frac{1}{n(m + 1)} \left( 1_{\{V_{M_i} = 1, V_{M_j} = 0, N = m+1\}} + 1_{\{V_{M_i} = 0, V_{M_j} = 1, N = m\}} \right).$$

For instance, for the first term there will be $n$ choices for $i$, and then $m$ choices for $j$ if the variable indexed by $i$ takes the value one and $N = m$, yielding $m$ zeros from which to choose $j$, or if the variable indexed by $i$ takes the value zero and $N = m + 1$, yielding $m$ ones from which to choose $j$.

Hence, taking conditional expectation we have

$$E\left(V^* - V \mid F\right) = U_n + W_n,$$

where

$$U_n = U_{n,1} + U_{n,2}$$

and

$$W_n = W_{n,1} + W_{n,2}.$$

17
Calculating the expectation of $U_{n,1}$, consider that for $i \neq j$,

$$P(V_i = 0, V_j = 0, V_{mi} = 0, V_{mj} = 1, N = m + 1, M = m) = P(V_i = 0, V_j = 0, V_{mi} = 0, V_{mj} = 1, N = m + 1|M = m)P(M = m)$$

$$= \frac{1}{2}P(V_i = 0, V_j = 0, V_{mi} = 0, V_{mj} = 1, N = m + 1)$$

$$= \frac{1}{4}P(V_i = 0, V_j = 0, V_{mi} = 0, V_{mj} = 1|N = m + 1)$$

$$= \frac{1}{4}g_{1,1,n,s_{m,m},m},$$

using (33), that is, that the conditional distribution of $V$ given $N = m + 1$ is that of $X_{s_{m,m}}$, followed by (57). Considering the case $M = m + 1$ similarly and averaging we obtain

$$P(V_i = 0, V_j = 0, V_{Mi} = 1, V_{Mj} = 0, N = m) = \frac{1}{4}g_{1,1,n,s_{m,m},m}.$$

Hence, letting

$$t_m = (1, \ldots, m - 1, m, m + 2, \ldots, n),$$

by (58) we have

$$EU_{n,1} = \frac{2(n)_2}{nm} \frac{1}{4}g_{1,1,n,s_{m,m},m} = \frac{(n)_2}{nm} \frac{1}{8}(1 - \lambda_{n,2,t_m}) \frac{m(m + 1)}{(n)_2} = \frac{m + 1}{8n} (1 - \lambda_{n,2,t_m}).$$

Likewise, for $U_{n,2}$, we have that

$$P(V_i = 0, V_j = 0, V_{mi} = 0, V_{mj} = 1, N = m, M = m) = \frac{1}{4}P(V_i = 0, V_j = 0, V_{mi} = 0, V_{mj} = 1|N = m)$$

$$= \frac{1}{4}g_{1,1,n,s_{m+1,m+1},m+1},$$

so with

$$t_{m+1} = (1, \ldots, m - 1, m + 1, m + 2, \ldots, n),$$

after averaging over $M = m + 1$, which yields the same result, by (58) we have

$$EU_{n,2} = \frac{2(n)_2}{m(m + 1)} \frac{1}{4}g_{1,1,n,s_{m+1,m+1},m} = \frac{m}{8n} (1 - \lambda_{n,2,t_{m+1}}).$$

For $W_{n,1}$, on $M = m$ we have

$$P(V_i = 0, V_j = 0, V_{mi} = 1, V_{mj} = 0, N = m, M = m) = \frac{1}{4}g_{1,1,n,s_{m+1,m+1},m},$$

with the same result on $M = m + 1$, hence

$$EW_{n,1} = \frac{2(n)_2}{nm} \frac{1}{4}g_{1,1,n,s_{m+1,m+1},m+1} = \frac{m + 1}{8n} (1 - \lambda_{n,2,t_{m+1}})$$

and likewise

$$EW_{n,2} = \frac{2(n)_2}{n(m + 1)} \frac{1}{4}g_{1,1,n,s_{m,m},m+1} = \frac{m}{8n} (1 - \lambda_{n,2,t_m}).$$

Summing all four contributions,

$$E(U_n + W_n) = \frac{1}{4} \left( \frac{\lambda_{n,2,t_m} + \lambda_{n,2,t_{m+1}}}{2} \right).$$
and so
\[
[E(U_n + W_n)]^2 = \frac{1}{64} \left( \lambda_{n,2,t_m}^2 + 2\lambda_{n,2,t_m}\lambda_{n,2,t_{m+1}} + \lambda_{n,2,t_{m+1}}^2 \right) - \frac{1}{16} (\lambda_{n,2,t_m} + \lambda_{n,2,t_{m+1}}) + \frac{1}{16}. \tag{38}
\]

To obtain the second moment of \(U_n + W_n\), note that
\[
(U_n + W_n)^2 = U_{n,1}^2 + U_{n,2}^2 + W_{n,1}^2 + W_{n,2}^2
\]
as the summands that define \(U_{n,1}, U_{n,2}, W_{n,1}\) and \(W_{n,2}\) are indicators of disjoint events. To calculate the expectation of \(U_{n,1}^2\), as in the proof of Lemma 3.3, we may write
\[
U_{n,1}^2 = U_{n,1,2}^2 + U_{n,1,3}^2 + U_{n,1,4}^2
\]
where
\[
U_{n,1,p}^2 = \frac{4}{n^2m^2} \sum_{\text{all four indices are distinct}} 1\{V_1 = 0, V_2 = 0, V_{M_1} = 0, V_{M_2} = 1, V_{t_1} = 0, V_{t_2} = 0, V_{M_{t_1}} = 0, V_{M_{t_2}} = 1, N = m+1\},
\]
with analogous definitions for \(U_{n,2,p}\) and a parallel decomposition for \(W_{n}\).

When all four indices are distinct, calculating using (57) and (59), as for the mean above, we find
\[
EU_{n,1,4}^2 = \frac{4(n_3^2 - 1)}{n^2m^2} g_{2,1,n,s_m,s_m,m} = \frac{m^2 - 1}{16n^2} (1 - 2\lambda_{n,2,t_m} + \lambda_{n,4,t_m}). \tag{40}
\]

With the inequalities over the summation in (39) in force, the event \(|\{i_1, i_2, j_1, j_2\}| = 3\) can occur in only the following two ways,

\[a) \ i_1 \neq i_2, j_1 = j_2 \quad \text{and} \quad b) \ i_1 = i_2, j_1 \neq j_2.\]

It is straightforward to see that the contribution from case \(a)\) is
\[
\frac{4(n_3^2 - 1)}{n^2m^2} g_{2,1,n,s_m,s_m,m} = \frac{m^2 - 1}{8n^2} (1 + \lambda_{n,1,t_m} - \lambda_{n,2,t_m} - \lambda_{n,3,t_m}),
\]
and from case \(b)\) is
\[
\frac{4(n_3^2 - 1)}{n^2m^2} g_{1,2,n,s_m,s_m,m} = \frac{m^2 - 1}{8n^2} (1 - \lambda_{n,1,t_m} - \lambda_{n,2,t_m} + \lambda_{n,3,t_m}),
\]
yielding
\[
EU_{n,1,3}^2 = \frac{m^2 + 1}{8n^2} (1 + \lambda_{n,1,t_m} - \lambda_{n,2,t_m} - \lambda_{n,3,t_m}) + \frac{m^2 - 1}{8n^2} (1 - \lambda_{n,1,t_m} - \lambda_{n,2,t_m} + \lambda_{n,3,t_m}).
\]

In view of the inequalities in (39), \(|\{i_1, i_2, j_1, j_2\}| = 2\) can only occur when \(i_1 = i_2\) and \(j_1 = j_2\), yielding
\[
EU_{n,1,2}^2 = \frac{4(n_3^2 - 1)}{n^2m^2} g_{1,1,n,s_m,s_m,m} = \frac{m^2 + 1}{4n^2} (1 - \lambda_{n,2,t_m}).
\]

In a completely analogous manner, we find that
\[
EU_{n,2,4}^2 = \frac{4(n_4^2 - 1)}{n^2(m+1)^2} g_{2,2,n,s_{m+1},s_{m+1},m+1} = \frac{m^2(m+1)}{16n^2(m+1)} (1 - 2\lambda_{n,2,t_{m+1}} + \lambda_{n,4,t_{m+1}}), \tag{41}
\]
\[
EU_{n,2,3}^2 = \frac{m^2}{8n^2(m+1)} (1 + \lambda_{n,1,t_{m+1}} - \lambda_{n,2,t_{m+1}} - \lambda_{n,3,t_{m+1}}) + \frac{m(m+1)}{8n^2(m+1)} (1 - \lambda_{n,1,t_{m+1}} - \lambda_{n,2,t_{m+1}} + \lambda_{n,3,t_{m+1}}),
\]

19
and that

\[ EU_{n,2,2}^2 = \frac{m}{4n^2(m + 1)}(1 - \lambda_{n,2,t_{m+1}}). \]

A parallel analysis yields

\[ EW_{n,1,4}^2 = \frac{m^2 - 1}{16n^2} \left( 1 - 2\lambda_{n,2,t_{m+1}} + \lambda_{n,4,t_{m+1}} \right), \]

(42)

\[ EW_{n,1,3}^2 = \frac{m + 1}{8n^2} \left( 1 + \lambda_{n,1,t_{m+1}} - \lambda_{n,2,t_{m+1}} - \lambda_{n,3,t_{m+1}} \right) + \frac{m^2 - 1}{8n^2m} \left( 1 - \lambda_{n,1,t_{m+1}} - \lambda_{n,2,t_{m+1}} + \lambda_{n,3,t_{m+1}} \right), \]

and that

\[ EW_{n,2,4}^2 = \frac{m^2(m - 1)}{16n^2(m + 1)} \left( 1 - 2\lambda_{n,2,t_{m}} + \lambda_{n,4,t_{m}} \right), \]

(43)

\[ EW_{n,2,3}^2 = \frac{m^2}{8n^2(m + 1)} \left( 1 + \lambda_{n,1,t_{m}} - \lambda_{n,2,t_{m}} - \lambda_{n,3,t_{m}} \right) + \frac{m(m - 1)}{8n^2(m + 1)} \left( 1 - \lambda_{n,1,t_{m}} - \lambda_{n,2,t_{m}} + \lambda_{n,3,t_{m}} \right), \]

and, lastly, that

\[ EW_{n,2,2}^2 = \frac{m}{4n^2(m + 1)}(1 - \lambda_{n,2,t_{m}}). \]

Collecting terms we obtain

\[
E(U_n + W_n)^2
= \frac{2m^3 + 6m^2 + 5m + 2}{8mn^2}
+ \frac{2m^2 + 2m + 1}{8mn^2(m + 1)}(\lambda_{n,1,t_{m}} - \lambda_{n,3,t_{m}} + \lambda_{n,1,t_{m+1}} - \lambda_{n,3,t_{m+1}})
- \frac{(m^2 + m + 1)(2m^2 + 2m + 1)}{8mn(m + 1)n^2}(\lambda_{n,2,t_{m}} + \lambda_{n,2,t_{m+1}}) + \frac{2m^3 - m - 1}{16n^2(m + 1)}(\lambda_{n,4,t_{m}} + \lambda_{n,4,t_{m+1}}). \]

Assuming without further mention in the remainder of the proof that \( n \geq 7 \), computing the difference of \( E(U_n + W_n)^2 - [E(U_n + W_n)]^2 \), with the squared expectation given in (38), firstly we have the constant term

\[
\frac{2m^3 + 6m^2 + 5m + 2}{8m(2m + 1)^2} - \frac{1}{16} = \frac{8m^2 + 9m + 4}{16m(2m + 1)^2} = \frac{4m(2m + 1) + 5m + 4}{16m(2m + 1)^2} = \frac{1}{4n} + \frac{5}{16n^2} + \frac{1}{2(n - 1)n^2}.
\]

Except for one mixed term, we only consider bounds for the terms involving \( t_m \), as the same bounds result for \( t_{m+1} \). For the coefficients of \( \lambda_{n,1,t_{m}} \) and \( \lambda_{n,3,t_{m}} \), noting these variables do not appear in (38), the inequality

\[
\frac{2m^2 + 2m + 1}{8mn^2(m + 1)} \leq \frac{1}{3n^2}
\]

and Lemma 5.3 yield the contributions

\[
\frac{2m^2 + 2m + 1}{8mn^2(m + 1)} |\lambda_{n,1,t_{m}}| \leq \frac{1}{3n^2} e^{1-n/2} \leq \frac{1}{m^2} e^{-n/2} \quad \text{and} \quad \left| \frac{2m^2 + 2m + 1}{8mn^2(m + 1)} |\lambda_{n,3,t_{m}}| \right| \leq \frac{1}{6n^2} e^{-n}.
\]
For the contribution from \( \lambda_{n,2,t_m} \), using Lemma 5.3,
\[
\left| -\frac{(m^2+m+1)(2m^2+2m+1)}{8m(m+1)n^2} + \frac{1}{16} \right| \lambda_{n,2,t_m} \leq \left( \frac{5}{16n^2} + \frac{5m+2}{16m(m+1)n^2} \right) |\lambda_{n,2,t_m}| \leq \frac{3e^{-n}}{8n^2}.
\]

Lastly we bound the difference
\[
\frac{2m^3-m-1}{16n^2(m+1)} \lambda_{n,4,t_m} - \frac{1}{64} \left( \lambda_{n,2,t_m}^2 + \lambda_{n,2,t_m} \lambda_{n,2,t_{m+1}} \right)
\]
\[
= \left[ \frac{2m^3-m-1}{16n^2(m+1)} - \frac{1}{32} \right] \lambda_{n,4,t_m} + \frac{1}{64} \left( 2\lambda_{n,4,t_m} - \lambda_{n,2,t_m}^2 - \lambda_{n,2,t_m} \lambda_{n,2,t_{m+1}} \right)
\]
\[
\leq \frac{8m^2+7m+3}{32(m+1)(2m+1)^2} \frac{1}{2} e^{-n} + \frac{1}{64} \left( e^{-n} + 2e^{-2n} \right)
\]
\[
= \left( \frac{(8m^2+4m)+3(m+1)}{32(m+1)(2m+1)^2} \right) \frac{1}{2} e^{-n} + \frac{1}{64} \left( e^{-n} + 2e^{-2n} \right)
\]
\[
\leq \left( \frac{1}{16n} + \frac{3}{64n^2} \right) e^{-n} + \frac{1}{64} \left( e^{-n} + 2e^{-2n} \right)
\]
\[
\leq \frac{1}{8n} e^{-n} + \frac{1}{64} e^{-n} + \frac{1}{32} e^{-2n}.
\]

Now adding in the corresponding terms involving \( t_{m+1} \), we conclude that the variance \( \text{Var}(U_n + W_n) \) is bounded by
\[
\frac{1}{4n} + \frac{5}{16n^2} + \frac{2}{2(n-1)n^2} + 2 \left( \frac{1}{n^2} e^{-n/2} + \frac{1}{6n^2} e^{-n} + \frac{3}{8n^2} e^{-n} + \frac{1}{8n} e^{-n} + \frac{1}{64} e^{-n} + \frac{1}{32} e^{-2n} \right)
\]
\[
\leq \frac{1}{4n} + \frac{5}{16n^2} + \frac{2}{2(n-1)n^2} + \frac{2}{n^2} e^{-n/2} + \left( \frac{1}{32} + \frac{1}{4n} + \frac{13}{12n^2} + \frac{1}{16} \right) e^{-n}.
\]

Now the upper bound (6) follows from the inequality \( \sqrt{a+b} \leq \sqrt{a} + \sqrt{b} \) for any \( a, b \geq 0 \). \( \square \)

We now provide a bound for the normal approximation of \( X \) in the odd case. We remark that fewer error terms, and therefore a smaller bound, results when standardizing \( X \) as in Theorem 1.1, that is, not by its own mean and variance but by the (exponentially close) mean and variance of the closely coupled \( V \).

**Proof of Theorem 1.1: odd \( n \).** Letting \( W = (X - \mu_V)/\sigma_V \) and \( V = (V - \mu_V)/\sigma_V \), since \( |X - V| \leq 1 \) by (34) of Theorem 4.1, we have
\[
|W - W_V| = |X - V|/\sigma_V \leq 1/\sigma_V.
\]

Letting \( P(Z \leq z) = \Phi(z) \) and
\[
C_V = \frac{\mu_V}{\sigma_V^2} \Delta_1 + 3.28 \frac{\mu_V}{\sigma_V^2} + \frac{2}{\sigma_V}.
\]

Using (44) and Theorem 4.2 we have
\[
P(W \leq z) - \Phi(z) \leq P(W_V - 1/\sigma_V \leq z) - \Phi(z)
\]
\[
= P(W_V \leq z + 1/\sigma_V) - \Phi(z + 1/\sigma_V) + \Phi(z + 1/\sigma_V) - \Phi(z)
\]
\[
\leq C_V + 1/(\sigma_V \sqrt{2\pi}).
\]

Similarly,
\[
P(W \leq z) - \Phi(z) \geq P(W_V + 1/\sigma_V \leq z) - \Phi(z)
\]
\[
= P(W_V \leq z - 1/\sigma_V) - \Phi(z - 1/\sigma_V) + \Phi(z - 1/\sigma_V) - \Phi(z)
\]
\[
\leq C_V + 1/(\sigma_V \sqrt{2\pi}),
\]
thus demonstrating the claim. \( \square \)
Appendix: Spectral Decomposition

In [10] the lightbulb chain was analyzed as a composition chain of multinomial type. Such chains in general are based on a $d \times d$ Markov transition matrix $P$ which describes the transition of a single particle in a system of $n$ identical particles, a subset of which is selected uniformly to undergo transition at each time step according to $P$.

In the case of the lightbulb chain there are $d = 2$ states and the transition matrix $P$ of a single bulb is given by

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

where we let $e_1 = (0, 1)^t$ and $e_0 = (1, 0)^t$ denote the 1 and 0 states of the bulb, for on and off, respectively. With $b \in \{0, 1, \ldots, n\}$ let $P_{n,b,s}$ be the $2^b \times 2^b$ transition matrix of a subset of size $b$ of the $n$ total lightbulbs when $s$ of the $n$ bulbs are selected uniformly to be switched. Letting $P_{n,0,s} = 1$ for all $n$ and $s$, and $I_2$ the $2 \times 2$ identity matrix, for $n \geq 1$ the matrix $P_{n,b,s}$ is given recursively by

$$P_{n,b,s} = \frac{s}{n} (P \otimes P_{n-1,b-1,s-1}) + (1 - \frac{s}{n}) (I_2 \otimes P_{n-1,b-1,s}) \text{ for } b \in \{1, \ldots, n\},$$

as any particular bulb among the $b$ in the subset considered is selected with probability $s/n$ to undergo transition according to $P$, leaving the $s-1$ remaining switches to be distributed over the remaining $b-1$ of $n-1$ bulbs, and with probability $1-s/n$ the bulb is left unchanged, leaving all the $s$ switches to be distributed.

The transition matrix $P$ is easily diagonalizable as

$$P = T^{-1} \Gamma T \quad \text{where} \quad T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad \Gamma = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

hence $P_{n,b,s}$ is diagonalized by

$$P_{n,b,s} = \otimes^b T^{-1} \Gamma_{n,b,s} \otimes^b T$$

(45)

where $\Gamma_{n,0,s} = 1$, and is otherwise given by the recursion

$$\Gamma_{n,b,s} = \frac{s}{n} (\Gamma \otimes \Gamma_{n-1,b-1,s-1}) + (1 - \frac{s}{n}) (I_2 \otimes \Gamma_{n-1,b-1,s}) \quad \text{for } b \in \{1, \ldots, n\}. \quad (46)$$

The next result describes the matrices $\Gamma_{n,b,s}$ more explicitly in terms of a sequence defined through the recursion

$$a_b = (a_{b-1}, a_{b-1} + 1_b) \quad \text{for } b \geq 2, \text{ with } a_1 = 0, \quad (47)$$

where $1_b = (1, \ldots, 1)$, a vector of length $2^b$. For example,

$$a_1 = (0, 1), \quad a_2 = (0, 1, 1, 2) \quad \text{and} \quad a_3 = (0, 1, 1, 2, 1, 2, 2, 3).$$

Letting $a_n$ be the $n^{th}$ term of the vector $a_b$ for any $b$ satisfying $2^b \geq n$ results in a well defined sequence $a_1, a_2, \ldots$

Lemma 5.1 For $n \in \{0, 1, \ldots, \}$ and $b, s \in \{0, \ldots, n\}$, and $\lambda_{n,b,s}$ given by (1), we have

$$\Gamma_{n,b,s} = diag(\lambda_{n,a_1,s}, \ldots, \lambda_{n,a_s,s}). \quad (48)$$

In particular, with $0_{2^{b-1}}$ the vector of zeros of length $2^{b-1}$, for $b \geq 1$

$$\Gamma_{n,b,s} = diag(\lambda_{n,a_1,s}, \ldots, \lambda_{n,a_{2^{b-1}-1},s}, 0_{2^{b-1}}) + diag(0_{2^{b-1}}, \lambda_{n,a_1+1,s}, \ldots, \lambda_{n,a_{2^{b-1}+1},s}).$$

22
For instance, 

\[ \Gamma_{n,2,s} = \text{diag}(\lambda_{n,0,s}, \lambda_{n,1,s}, \lambda_{n,1,s}, \lambda_{n,2,s}), \]

and 

\[ \Gamma_{n,3,s} = \text{diag}(\lambda_{n,0,s}, \lambda_{n,1,s}, \lambda_{n,1,s}, \lambda_{n,2,s}, \lambda_{n,2,s}, \lambda_{n,3,s}). \]

**Proof.** As \( \lambda_{n,0,s} = 1 \), the lemma is true for \( b = 0 \). For the inductive step, assuming the lemma is true for \( b - 1 \), by (46) it suffices to verify

\[
- \frac{s}{n} \lambda_{n-1,a,s-1} + (1 - \frac{s}{n}) \lambda_{n-1,a,s} = \lambda_{n,a,s} \quad \text{and} \\
- \frac{s}{n} \lambda_{n-1,a,s-1} + (1 - \frac{s}{n}) \lambda_{n-1,a,s} = \lambda_{n,a+1,s}.
\]

For the first equality, by (1), the left hand side equals

\[
\frac{s}{n} \sum_{t=0}^{a} \binom{a}{t} (-2)^t \frac{(s-1)}{(n-1)t} + (1 - \frac{s}{n}) \sum_{t=0}^{a} \binom{a}{t} (-2)^t \frac{s}{(n-1)t} \\
= \sum_{t=0}^{a} \binom{a}{t} (-2)^t \left( \frac{s}{n} \frac{(s-1)}{(n-1)t} + (1 - \frac{s}{n}) \frac{s}{(n-1)t} \right) \\
= \sum_{t=0}^{a} \binom{a}{t} (-2)^t \frac{(s+1) + (n-s)(s)}{(n+1)t} \\
= \sum_{t=0}^{a} \binom{a}{t} (-2)^t \frac{(s+1) + (n-s)}{(n+1)t} \\
= \frac{\lambda_{n,a,s}}{n}.
\]

The second equality can be shown in similar, though slightly more involved, fashion. \( \square \)

We note that [10] expresses these eigenvalues in terms of the hypergeometric function.

If the \( k \) stages of the process \( 1, \ldots, k \) use switches \( s = (s_1, \ldots, s_k) \), as (45) implies that the matrices \( P_{n,b,s}, s \in \{0, 1, \ldots, n\} \) are simultaneously diagonalizable, the overall transition matrix \( P_{n,b,s} \) for a subset of \( b \) bulbs is the product

\[
\prod_{j=1}^{k} P_{n,b,s_j} = \otimes^b T^{-1} \Gamma_{n,b,s} \otimes^b T = \otimes^b T^{-1} \text{diag}(\lambda_{n,a_1,s}, \ldots, \lambda_{n,a_{kb},s}) \otimes^b T, \tag{48}
\]

where

\[
\Gamma_{n,b,s} = \prod_{j=1}^{k} \Gamma_{n,b,s_j} \quad \text{and} \quad \lambda_{n,a,s} = \prod_{j=1}^{k} \lambda_{n,a,s_j}. \tag{49}
\]

Hence, if \( \pi \) is any permutation of \( \{1, \ldots, k\} \), letting \( \pi(s) = (s_{\pi(1)}, \ldots, s_{\pi(k)}) \), from (49) we have \( \Gamma_{n,b,s} = \Gamma_{n,b,\pi(s)} \), and now from (48) that

\[
P_{n,b,s} = P_{n,b,\pi(s)}. \tag{50}
\]
As order is unimportant, we may replace \((s_1, \ldots, s_k)\) by the multiset \(\{s_1, \ldots, s_k\}\) when convenient.

The following lemma computes the probabilities that out of a group of \(2r\) bulbs, starting with initial conditions such as half the bulbs off and the other half on, after \(k\) transitions using \(s_1, \ldots, s_k\) switches, all bulbs will be off.

**Lemma 5.2** With \(s = (s_1, \ldots, s_k)\) let \(\lambda_{n,a,s}\) be given by (49). Then for any \(r \in \{0, 1, \ldots\}\) with \(2r \leq n\),

\[
P(X_i = 0, i = 1, \ldots, 2r | X_0, i \equiv i \mod 2, i = 1, \ldots, 2r) = \frac{1}{2^{2r}} \sum_{j=0}^r (-1)^j \binom{r}{j} \lambda_{n,2j,s},
\]

and for any \(r \in \{0, 1, \ldots, n\}\),

\[
P(X_i = 0, i = 1, \ldots, r | X_0, i = 0, i = 1, \ldots, r-1, X_{0,r} = 1) = \frac{1}{2^r} \sum_{j=0}^r (1 - 2j/r) \binom{r}{j} \lambda_{n,j,s}, \tag{51}
\]

and

\[
P(X_i = 0, i = 1, \ldots, r | X_0, i = 1, i = 1, \ldots, r-1, X_{0,r} = 0) = \frac{1}{2^r} \sum_{j=0}^r (-1)^j (1 - 2j/r) \binom{r}{j} \lambda_{n,j,s}. \tag{52}
\]

**Proof.** Letting \(e_1 = (0, 1)^t\) and \(e_0 = (1, 0)^t\), extracting the appropriate element of the transition matrix, we have

\[
P(X_i = 0, i = 1, \ldots, 2r | X_0, i \equiv i \mod 2, i = 1, \ldots, 2r) = (e_1 \otimes e_0)^\otimes r P_{n,2r,s} \otimes 2r.
\]

For \(j \in \{0, 1, \ldots\}\) let \(\Omega_{2r,j}\) be the \(2^{2r} \times 2^{2r}\) diagonal matrix in the variables \(x_k, k \in \{0, 1, \ldots\}\) given by

\[
\Omega_{2r,j} = \text{diag}(x_{a_1+j}, \ldots, x_{a_{2r}+j}),
\]

and

\[
w_{2r} = (e_1^t \otimes e_0) \otimes 2^r T^{-1} \quad \text{and} \quad u_{2r} = \otimes 2^r T e_0 \otimes 2^r.
\]

With \(\Omega_{2r} = \Omega_{2r,0}\) we verify by induction that

\[
w_{2r}, \Omega_{2r}, u_{2r} = \sum_{j=0}^r (-1)^j \binom{r}{j} x_{2j}. \tag{53}
\]

Adopting the convention that \(a^{\otimes 0} = 1\), sides of (53) equal \(x_0\) when \(r = 0\), so we assume (53) holds for nonnegative integers less than \(r\). As \(e_1^t T^{-1} = (1, 1)/\sqrt{2}\), and \(e_0^t T^{-1} = (1, -1)/\sqrt{2}\) we have

\[
w_{2r} = (e_1^t \otimes e_0) \otimes 2^{T^{-1}} \otimes 2^r w_{2r-2} = \frac{1}{2}(w_{2r-2}, -w_{2r-2}, w_{2r-2}, -w_{2r-2}), \tag{54}
\]

and similarly, as \(T e_0 = (1, -1)^t/\sqrt{2}\)

\[
u_{2r} = \frac{1}{2}(u_{2r-2}, -u_{2r-2}, -u_{2r-2}, u_{2r-2}). \tag{55}
\]

Using (47), we may write

\[
\Omega_{2r} = \begin{bmatrix} \Omega_{2r-1,0} & 0 \\ 0 & \Omega_{2r-1,1} \end{bmatrix} = \begin{bmatrix} \Omega_{2r-2,0} & 0 & 0 & 0 \\ 0 & \Omega_{2r-2,1} & 0 & 0 \\ 0 & 0 & \Omega_{2r-2,1} & 0 \\ 0 & 0 & 0 & \Omega_{2r-2,2} \end{bmatrix}.
\]
Now calculating the left hand side of (53) using (54), (55), and (56) yields
\[
\frac{1}{4} \left( w_{2r-2} \Omega_{2r-2,0} u_{2r-2} + w_{2r-2} \Omega_{2r-2,1} u_{2r-2} - w_{2r-2} \Omega_{2r-2,1} u_{2r-2} - w_{2r-2} \Omega_{2r-2,2} u_{2r-2} \right)
= \frac{1}{4} \left( w_{2r-2} \Omega_{2r-2,0} u_{2r-2} - w_{2r-2} \Omega_{2r-2,2} u_{2r-2} \right).
\]

By the induction hypotheses, the first term is given by (53) with \( r \) replaced by \( r - 1 \), and the second term the same as the first term, but with \( x_{2j} \) replaced by \( x_{2j+2} \). Hence, we obtain
\[
w_{2r} \Omega_{2r,u_{2r}} = \frac{1}{4} \left( \frac{1}{2^{2(r-1)}} \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} x_{2j} - \frac{1}{2^{2(r-1)}} \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} x_{2j+2} \right)
= \frac{1}{2^{2r}} \left( \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} x_{2j} - \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} x_{2j+2} \right)
= \frac{1}{2^{2r}} \left( \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} x_{2j} + \sum_{j=1}^{r} (-1)^j \binom{r-1}{j-1} x_{2j} \right)
= \frac{1}{2^{2r}} \sum_{j=0}^{r} (-1)^j \binom{r}{j} x_{2j},
\]
as desired.

Now to prove the first claim of the lemma note that the diagonalization (45) yields
\[(e_0^l \otimes e_0^l)^{\otimes r} P_{n,2r,s_0}^{\otimes 2r} = w_{2r} \Gamma_{n,2r,s} u_{2r},\]
and apply (53). Equalities (51) and (52) can be shown in a similar, but simpler, manner. \( \square \)

Suppose switches \( s = (s_1, \ldots, s_n) \) are applied to \( n \) bulbs in stages 1, \ldots, \( n \), respectively. Then, for stage \( l \), with \( \alpha \) and \( \beta \) nonnegative with \( \alpha + \beta \leq n \),
\[P(X_{l1} = \cdots = X_{l\alpha} = 0, X_{l,\alpha+1} = \cdots = X_{l,\alpha+\beta} = 1) = \frac{(n-s_l)_{\alpha}(s_l)_\beta}{(n)_{\alpha+\beta}},\]
so we define
\[g_{\alpha,\beta,n,s,l} := P(X_{l1} = \cdots = X_{\alpha+\beta} = 0, X_{l1} = \cdots = X_{l\alpha} = 0, X_{l,\alpha+1} = \cdots = X_{l,\alpha+\beta} = 1) = P(X_{l1} = \cdots = X_{\alpha+\beta} = 0 | X_{l1} = \cdots = X_{l\alpha} = 0, X_{l,\alpha+1} = \cdots = X_{l,\alpha+\beta} = 1) \frac{(n-s_l)_{\alpha}(s_l)_\beta}{(n)_{\alpha+\beta}}.\]

By (50), that is, the fact that the switch variables can be applied in any order, the conditional probability in (57) is the same as that for the lightbulb process with initial condition \( X_{l1} = \cdots = X_{l\alpha} = 0, X_{l,\alpha+1} = \cdots = X_{l,\alpha+\beta} = 1 \) which skips stage \( l \). Hence, for given \( s \) and \( l \in \{1, \ldots, n\} \), letting \( s_l = (s_1, \ldots, s_{l-1}, s_{l+1}, \ldots, s_n) \),
by Lemma 5.2,
\[g_{r,r,n,s,l} = \frac{1}{2^{2r}} \sum_{j=0}^{r} (-1)^j \binom{r}{j} \lambda_{n,2j,s_l} \frac{(n-s_l)_r(s_l)_r}{(n)_{2r}}\]
and
\[g_{r-1,1,n,s,l} = \frac{1}{2^r} \sum_{j=0}^{r} (1 - 2j/r) \binom{r}{j} \lambda_{n,j,s_l} \frac{(n-s_l)_{r-1}(s_l)_1}{(n)_r}\]
\[g_{1,r-1,n,s,l} = \frac{1}{2^r} \sum_{j=0}^{r} (-1)^j (1 - 2j/r) \binom{r}{j} \lambda_{n,j,s_l} \frac{(n-s_l)(s_l)_{r-1}}{(n)_r}.\]
Specializing further, we obtain
\[ g_{1, n, s, t} = \frac{1}{4} (1 - \lambda_{n, 2, s_t}) \frac{s_t(n - s_t)}{(n)_2}, \]  
(58)
\[ g_{2, n, s, t} = \frac{1}{16} (1 - 2\lambda_{n, 2, s_t} + \lambda_{n, 4, s_t}) \frac{(s_t)^2(n - s_t)^2}{(n)_4}, \]  
(59)
\[ g_{2, n, s, t} = \frac{1}{8} (1 + \lambda_{n, 1, s_t} - \lambda_{n, 2, s_t} - \lambda_{n, 3, s_t}) \frac{(n - s_t)^2 s_t}{(n)_3}, \]  
(60)
and
\[ g_{1, n, s, t} = \frac{1}{8} (1 - \lambda_{n, 1, s_t} - \lambda_{n, 2, s_t} + \lambda_{n, 3, s_t}) \frac{(n - s_t)(s_t)^2}{(n)_3}. \]  
(61)

In order to obtain bounds on
\[ \Delta = \sqrt{\text{Var}(E(X^* - X|X))} \]
as required by Theorem 1.1, we study products of the eigenvalues of the chain.

**Lemma 5.3** For all even \( n \geq 6 \) and \( s = (1, \ldots, n/2 - 1, n/2 + 1, \ldots, n) \),
\[ |\lambda_{n, 2, s}| \leq e^{-n} \quad \text{and} \quad |\lambda_{n, 4, s}| \leq \frac{1}{2} e^{-n}. \]  
(62)

For all odd \( n = 2m + 1 \geq 7 \) and \( s \) equal to
\[ t_m = (1, \ldots, m - 1, m + 1, \ldots, n) \quad \text{or} \quad t_{m+1} = (1, \ldots, m - 1, m + 2, \ldots, n), \]
both inequalities in (62) hold, as do
\[ |\lambda_{n, 1, s}| \leq e^{-n/2} \quad \text{and} \quad |\lambda_{n, 3, s}| \leq \frac{1}{2} e^{-n}. \]  
(63)

**Proof.** In the even case we define \( m \) by \( n = 2m \), and in the odd case we consider only \( t_m \), the argument for \( t_{m+1} \) being essentially identical. We follow and slightly generalize the arguments of [6].

Starting with the first claim in (62), for \( n \geq 2 \) consider the second degree polynomial
\[ f_2(x) = 1 - \frac{4x}{n} + \frac{4(x)^2}{(n)_2}, \quad 0 \leq x \leq n. \]
It is simple to verify that \( f_2(x) \) achieves its global minimum value of \(-1/(n - 1)\) at \( n/2 \), and that \( f_2(x) \) has exactly two roots, at \((n + \sqrt{n})/2\) and \((n - \sqrt{n})/2\). Hence, as \( f_2(x) \leq 0 \) for all \( x \) between these roots, and additionally, as \((x - 1)/(n - 1) \leq x/n\) for all \( x \in [0, n] \), we obtain the bound
\[ |f_2(x)| \leq \left\{ \begin{array}{ll}
\frac{1}{n - 1} & \text{for } x \in \left[\frac{n - \sqrt{n}}{2}, \frac{n + \sqrt{n}}{2}\right]
\frac{1}{(1 - x)^2} & \text{for } x \in [0, n], |x - n/2| > \sqrt{n}/2.
\end{array} \right. \]  
(64)

For \( x \in \mathbb{R} \) let \([x]\) and \([x]\) denote the greatest integer less than or equal to \( x \), and the smallest integer greater than or equal to \( x \). Further, recalling that \( m \) is given by \( n = 2m \) and \( n = 2m + 1 \) in the even and odd cases, respectively, let
\[ t = \left\{ \left[\frac{n - \sqrt{n}}{2}\right], \ldots, \left[\frac{n + \sqrt{n}}{2}\right] \right\} \setminus \{m\} \]
Now, in both the even and odd cases we have

$$|\lambda_{n,2,s}| = \left( \prod_{s=0}^{[n-\sqrt{n}/2]} |f_2(s)| \right) \left( \prod_{s=t}^{[n]} |f_2(s)| \right) \left( \prod_{s=\lceil n+\sqrt{n}/2 \rceil}^{n} |f_2(s)| \right),$$

noting additionally that if either \((n-\sqrt{n})/2\) or \((n+\sqrt{n})/2\) is an integer then equality holds as both expressions above are zero. We may henceforth assume neither value is an integer, with similar remarks applying to the products (68) and (69).

Applying the bound (64), and that \(|n/2 - x| + |n/2 + x| = n\), yields

$$|\lambda_{n,2,s}| \leq \left( \prod_{s=0}^{[n-\sqrt{n}/2]} \left( 1 - \frac{2s}{n} \right)^2 \right) \left( \prod_{s=t}^{[n]} \frac{1}{n-1} \right) \left( \prod_{s=\lceil n+\sqrt{n}/2 \rceil}^{n} \left( 1 - \frac{2s}{n} \right)^2 \right)$$

$$= \left( \prod_{s=0}^{[n-\sqrt{n}/2]} \left( 1 - \frac{2s}{n} \right) \right)^4 \left( \frac{1}{n-1} \right)^{|t|},$$

where \(|t|\) is the cardinality of \(t\).

Using \(1 - x \leq e^{-x}\) for \(x \geq 0\) and that \(|x| \geq x - 1\) on the first product, we obtain the bound

$$|\lambda_{n,2,s}| \leq \left( e^{-\frac{2}{n}([n-\sqrt{n}/2]) \left( \frac{2}{n}([n-\sqrt{n}/2]) + 1 \right)/2} \right)^4 e^{-|t| \log(n-1)} \leq e^{-(n-2\sqrt{n}+2/\sqrt{n}+|t| \log(n-1))}. \quad (65)$$

To control \(|t|\), note that as \(|x| \leq x + 1\), we have

$$|t| = \left\lfloor \frac{n + \sqrt{n}}{2} \right\rfloor - \left\lfloor \frac{n - \sqrt{n}}{2} \right\rfloor \geq \sqrt{n} - 2.$$

As \(\log 35 \geq 3.5\), for \(n \geq 36\) we have

$$-2\sqrt{n} - 1 + 2/\sqrt{n} + |t| \log(n-1) \geq -2\sqrt{n} - 1 + 3.5(\sqrt{n} - 2) = \frac{3}{2}\sqrt{n} - 8 \geq 0,$$

and hence, from (65), that

$$|\lambda_{n,2,s}| \leq e^{-n} \quad \text{for } n \geq 36.$$

Using the given choices for \(s\) in the even and odd cases, one may verify directly that \(\lambda_{n,2,s}\) satisfies this same bound for all even integers \(6 \leq n \leq 34\), and all odd integers \(7 \leq n \leq 35\), completing the proof for all claims on \(\lambda_{n,2,s}\).

Now turning to \(\lambda_{n,4,s}\), for \(n \geq 4\) consider the fourth degree polynomial

$$f_4(x) = 1 - \frac{8x}{n} + \frac{24(x)^2}{(n)^2} - \frac{32(x)^3}{(n)^3} + \frac{16(x)^4}{(n)^4}, \quad 0 \leq x \leq n.$$ 

It can be checked that the four roots of \(f_4(x)\) are given by

$$x_{1\pm} = \frac{n \pm \sqrt{2/\sqrt{3n^2 - 9n + 8 + 3n - 4}}}{2} \quad \text{and} \quad x_{2\pm} = \frac{n \pm \sqrt{2/\sqrt{3n^2 - 9n + 8 + 3n - 4}}}{2},$$

and that additionally the three roots to the cubic equation \(f'_4(x) = 0\) occur at

$$y_1 = n/2 \quad \text{and} \quad y_{2\pm} = (n \pm \sqrt{3n-4})/2.$$
These roots satisfy
\[ 0 < y_1 < y_2 < y_3 < y_4 < y_5 < y_6 < x_1 < x_2. \]

To obtain a bound over the interval \([x_1, x_1+1]\), evaluating \(f_4(x)\) at its critical values we obtain
\[
\begin{align*}
f_4(y_1) &= \frac{3}{(n-1)(n-3)} \leq \frac{3}{(n-3)^2} \quad \text{and} \\
f_4(y_{2\pm}) &= -\frac{6(2n^2 - 9n + 8)}{n(n-1)(n-2)(n-3)} \geq -\frac{6}{(n-3)^2}.
\end{align*}
\]

To bound \(f_4(x)\) by \(f_2^2(x)\) in the remaining part of \([0, n]\), write
\[
\begin{align*}
f_2^2(x) - f_4(x) &= \frac{16(n-x)p(x)}{(n-1)^2n^2(n^2 - 5n + 6)}
\end{align*}
\]
where
\[
p(x) = (4n-6)x^2 + (6n-4n^2)x + n^3 - 2n^2 + n.
\]
The roots of the quadratic \(p(x)\) are given by
\[
z_{\pm} = \frac{n}{2} \pm \frac{1}{2} \sqrt{n(n-2)}.
\]
As \(5n^2 - 15n + 12 \geq 0\) for all \(n\), we have \((2n-3)(3n-4) \geq n(n-2)\), and therefore
\[
(\sqrt{2}\sqrt{3n^2 - 9n + 8} + 3n - 4)(2n-3) \geq n(n-2).
\]
Dividing by \(2n-3\) and taking square roots demonstrates that
\[ x_1 - z_- < z_+ < x_1. \]
Hence \(p(x)\) is nonnegative on the complement of \([z_-, z_+]\), and we obtain
\[
|f_4(x)| \leq \begin{cases} 
\frac{6}{(n-3)^2} & \text{for } x \in [x_1, x_1+1] \\
\frac{6}{f_2^2(x)} & \text{for } x \notin [x_1, x_1+1], x \in [0, n].
\end{cases}
\]
Now write for short
\[
C(n) = \sqrt{2}\sqrt{3n^2 - 9n + 8} + 3n - 4 \quad \text{so that} \quad x_1 = \frac{n - C(n)}{2}.
\]
Using (66), (65) and that \(|s-n/2| > \sqrt{n}/2\) whenever \(s \leq (n - C(n))/2, 1 - x \leq e^{-x}\) for \(x \geq 0\) and finally \(x \geq x - 1\), we obtain
\[
\prod_{s=0}^{\lfloor n - C(n) \rfloor} |\lambda_{n,4,s}| \leq \prod_{s=0}^{\lfloor n - C(n) \rfloor} f_2^2(s) \leq \left( \prod_{s=0}^{\lfloor n - C(n) \rfloor} \left( 1 - \frac{2s}{n} \right) \right)^4 \leq e^{-2(n-C(n)+C(n)/n+2C(n)/n)}. \quad (67)
\]
Now let
\[
u = \left\{ \left\lfloor \frac{n - C(n)}{2} \right\rfloor, \ldots, \left\lfloor \frac{n + C(n)}{2} \right\rfloor \right\} \setminus \{m\}.
\]
Using $\lambda_{n,4,s} = \lambda_{n,4,n-s}$, (66), $|u| \geq C(n) - 2$ and (67), we have, in both the even and odd cases, that

$$
|\lambda_{n,4,s}| = \prod_{s=0}^{n-C(n)} |\lambda_{n,4,s}| \prod_{s=\left\lceil \frac{n-C(n)}{2} \right\rceil}^{n} |\lambda_{n,4,s}| \leq \left( \prod_{s=0}^{n-C(n)} \lambda_{n,4,s}^{2} \right) \left( \frac{6}{(n-3)^2} \right)^{C(n)-2} (68)
$$

Therefore,

$$
\prod_{s=m+2}^{2m+1} \left( 1 - \frac{2s}{n} \right) = (-1)^{m} \prod_{s=0}^{m-1} \left( 1 - \frac{2s}{n} \right) \quad \text{and hence} \quad \prod_{s=m+2}^{2m+1} \left| 1 - \frac{2s}{n} \right| \leq e^{-m(m-1)/n}.
$$

Thus,

$$
|\lambda_{n,1,s,m}| \leq e^{-2m^2/n} = e^{-(n-1)^2/(2n)} \leq e^{-(n/2 - 1)}.
$$

for all $n \geq 1$.

Now, to obtain the claimed bound for $\lambda_{n,3,s,m}$ consider the third degree polynomial

$$
f_{3}(x) = 1 - \frac{6x}{n} + \frac{12(x-1)}{n^2} - \frac{8(x-1)^3}{n^3}, \quad 0 \leq x \leq n.
$$

The three roots of $f_{3}(x)$ are given by

$$
x_{1\pm} = \frac{n \pm \sqrt{3n-2}}{2} \quad \text{and} \quad x_{2} = \frac{n}{2}.
$$

Additionally, $f_{3}(x)$ achieves its local extreme values at

$$
y_{1\pm} = \frac{3n \pm \sqrt{3} \sqrt{3n-2}}{6} \quad \text{for which} \quad |f_{3}(y_{1\pm})| = \frac{2(n-2/3)^{3/2}}{n(n-1)(n-2)} \leq \frac{2}{(n-2)^{3/2}}.
$$

Clearly $x_{1-} < y_{1-} < x_{2} < y_{1+} < x_{1+}$, so $|f_{3}(x)| \leq |f_{3}(y_{1\pm})|$ over $[x_{1-}, x_{1+}]$.

To bound $|f_{3}(x)|$ over the remaining portion of $[0, n]$, we have

$$
\left( 1 - \frac{2x}{n} \right)^3 - f_{3}(x) = \frac{4(3n-2)x(n-x)(n-1)}{n^3(n^2-3n+2)},
$$

29
demonstrating that \( f_3(x) \leq (1 - 2x/n)^3 \) for all \( x \in [0, n/2] \), and hence \( |f_3(x)| \leq |(1 - 2x/n)^3| \) for all \( x \in [0, x_1-] \). By the odd symmetry of \( f_3(x) \) and \( (1 - 2x/n) \) about \( x = n/2 \), we now obtain

\[
|f_3(x)| \leq \begin{cases} 
\frac{2}{(n-2)^{3/2}} & \text{for } x \in [x_1-, x_1+] \\
\left| 1 - \frac{2x}{n} \right|^3 & \text{for } x \not\in [x_1-, x_1+], x \in [0, n]. 
\end{cases}
\]

Hence, arguing as before,

\[
\prod_{s=0}^{[x_1-]} |\lambda_{n,3,s}| \leq \prod_{s=0}^{[x_1-]} \left( 1 - \frac{2s}{n} \right) \geq e^{-\frac{2}{3}(n-2\sqrt{3n-2}+(n-2)/n+2\sqrt{3n-2}/n)};
\]

with the same bound holding for \( \prod_{s=[x_1+]}^{n} |\lambda_{n,3,s}| \). Therefore, letting

\[
u = \left\{ \frac{n - \sqrt{3n-2}}{2}, \ldots, \frac{n + \sqrt{3n-2}}{2} \right\} \setminus \{ m \}
\]

we have \( |\nu| \geq \sqrt{3n-2} - 2 \) and

\[
|\lambda_{n,3,t_m}| = \prod_{s=1}^{[x_1-]} |\lambda_{n,3,s}| \prod_{s \in \nu} \prod_{s=[x_1+]}^{n} |\lambda_{n,3,s}| \leq \prod_{s=1}^{[x_1-]} \lambda_{n,3,s}^2 \left( \frac{2}{n-2} \right)^{\sqrt{3n-2}-2}
\]

\[
\leq e^{-\frac{2}{3}(n-2\sqrt{3n-2}+(n-2)/n+2\sqrt{3n-2}/n+(\sqrt{3n-2}-2)(\log(2)-\log(n-2)))}
\]

\[
= e^{-n} \cdot e^{-\left( \frac{1}{2}n - 3\sqrt{3n-2} + \frac{1}{2}(n-2)/n + 3\sqrt{3n-2}/n + (\sqrt{3n-2}-2)(\frac{1}{2}\log(n-2)-\log(2)) \right)}.
\]

Since for \( n \geq 111 \) we have

\[
\frac{1}{2}n - 3\sqrt{3n-2} \geq \log(2) \quad \text{and} \quad \frac{3}{2}\log(n-2) - \log(2) \geq 0,
\]

we obtain that \( |\lambda_{n,3,s}| \leq e^{-n/2} \) over this range. Verifying directly that the claims of the lemma hold for odd \( 7 \leq n \leq 111 \) completes the proof. \( \square \)

**Bibliography**

1. Baldi, P., Rinott, Y. and Stein, C. (1989). A normal approximation for the number of local maxima of a random function on a graph, In *Probability, Statistics and Mathematics, Papers in Honor of Samuel Karlin*. T. W. Anderson, K.B. Athreya and D. L. Iglehart eds., Academic Press, 59-81.

2. Gosh, S. and Goldstein, L. (2009). Concentration of measures via size biased couplings. *Probability Theory and Related Fields*, to appear

3. Goldstein, L. (2005). Berry-Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing. *J. Appl. Probab*. 42, 661–683.

4. Goldstein, L. and Rinott, Y. (1996). On multivariate normal approximations by Stein’s method and size bias couplings. *J. Appl. Prob*. 33, 1-17.

5. Goldstein, L. and Penrose, M. (2009). Normal approximation for coverage models over binomial point processes. *Ann. Appl. Prob*. to appear.

6. Rao, C., Rao, M., and Zhang, H. (2007). One Bulb? Two Bulbs? How many bulbs light up? A discrete probability problem involving dermal patches. *Sankhyā*, 69, pp. 137-161.
7. Shao, Q. M. and Su, Z. (2005). The Berry-Esseen bound for character ratios. *Proceedings of the American Mathematical Society* **134**, 2153-2159.

8. Stein, C. (1981) Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9**, 1135-1151.

9. Stein, C. (1986). Approximate Computation of Expectations. IMS, Hayward, California.

10. Zhou, H. and Lange, K. (2009). Composition Markov chains of multinomial type. *Advances in Applied Probability*, **41**, pp. 270 - 291.