

## A BERRY–ESSEEN BOUND FOR THE LIGHTBULB PROCESS

LARRY GOLDSTEIN,\* *University of Southern California*

HAIMENG ZHANG,\*\* *Mississippi State University*

### Abstract

In the so-called lightbulb process, on days  $r = 1, \dots, n$ , 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  $\sup_{z \in \mathbb{R}} |\mathbb{P}((X - \mu)/\sigma \leq z) - \mathbb{P}(Z \leq z)| \leq n\bar{\Delta}_0/2\sigma^2 + 1.64n/\sigma^3 + 2/\sigma$ , where  $Z$  is a standard normal random variable and  $\bar{\Delta}_0 = 1/2\sqrt{n} + 1/2n + e^{-n/2}/3$  for  $n \geq 6$ , yielding a bound of order  $O(n^{-1/2})$  as  $n \rightarrow \infty$ . A similar, though slightly larger bound, holds for odd  $n$ . 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  $X$  that has the  $X$ -size bias distribution, that is, which satisfies  $\mathbb{E}[Xg(X)] = \mu \mathbb{E}[g(X^s)]$  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 odd  $n$  is similar to that for even  $n$ , 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. In both the even and odd cases, the crucial calculation of the variance of a conditional expectation requires detailed information on the spectral decomposition of the lightbulb chain.

*Keywords:* Normal approximation; toggle switch; Stein’s method; size biasing

2010 Mathematics Subject Classification: Primary 62E17; 60C05  
Secondary 60B15; 62P10

### 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 each day of the study, some number of 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 from [8]. Consider  $n$  toggle switches, each connected to a lightbulb, all of which are initially off. 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 [8] for the case where on each day  $r = 1, \dots, n$ , exactly  $r$  of the  $n$  switches are randomly pressed.

Received 11 January 2010; revision received 21 February 2011.

\* Postal address: Department of Mathematics, University of Southern California, 1042 West 36th Place, KAP 108, Los Angeles, CA 90089-2532, USA. Email address: larry@math.usc.edu

\*\* Postal address: Department of Mathematics and Statistics, Mississippi State University, Mississippi State, MS 39762, USA.

More generally, consider the lightbulb process on  $n$  bulbs with some number  $k$  of stages, where  $s_r \in \{0, \dots, n\}$  lightbulbs are toggled in stage  $r$ , for  $r = 1, \dots, k$ ; we refer to the vector  $s = (s_1, \dots, s_k)$  recording the number of bulbs affected on each study day as the ‘switch pattern’. In order to consider quantities that depend on some subset of size  $b$  of the  $n$  bulbs, we define

$$\lambda_{n,b,s} = \sum_{t=0}^b \binom{b}{t} (-2)^t \frac{(s)_t}{(n)_t} \quad \text{and} \quad \lambda_{n,b,s} = \prod_{r=1}^k \lambda_{n,b,s_r}, \tag{1.1}$$

where  $(n)_k = n(n-1) \cdots (n-k+1)$  denotes the falling factorial, and the empty product is 1. Generalizing the results in [8], writing  $X_s$  for the number of bulbs on at the terminal time when applying the switch pattern  $s = (s_1, \dots, s_n)$ , the martingale method in Proposition 4 of [12] shows that if the process is initialized with all bulbs off then

$$E X_s = \frac{n}{2}(1 - \lambda_{n,1,s}) \quad \text{and} \quad \text{var}(X_s) = \frac{n}{4}(1 - \lambda_{n,2,s}) + \frac{n^2}{4}(\lambda_{n,2,s} - \lambda_{n,1,s}^2), \tag{1.2}$$

where, from (1.1),

$$\lambda_{n,1,s} = 1 - \frac{2s}{n} \quad \text{and} \quad \lambda_{n,2,s} = 1 - \frac{4s}{n} + \frac{4s(s-1)}{n(n-1)} \quad \text{for } s = 1, \dots, n. \tag{1.3}$$

Letting  $\mathbf{n} = (1, \dots, n)$ , we call the standard lightbulb process the one where  $s = \mathbf{n}$ , and in this case we will write  $X$  short for  $X_{\mathbf{n}}$ . In particular, (1.2) with  $s = \mathbf{n}$  recovers the mean  $\mu = E X$  and variance of  $\sigma^2 = \text{var}(X)$  as computed in [8]. Other results in [8] include recursions for determining the exact finite sample distribution of  $X$ . Though computational approximations to the distribution of  $X$ , including by the normal, were also considered in [8], the quality of such approximations, and the asymptotic normality of  $X$ , was left open.

Theorem 1.1 below settles the matter of the asymptotic distribution of  $X$  by providing a bound to the normal which holds for all finite  $n$ , and which tends to 0 at the rate  $n^{-1/2}$  as  $n$  tends to  $\infty$ . We consider the cases of even and odd  $n$  separately. In the even case we directly couple the variable  $X$  to a variable having the  $X$ -size bias distribution, as described later on in this section. In the even case (1.3) yields  $\lambda_{n,1,n/2} = 0$ , and, therefore,  $\lambda_{n,1,n} = 0$ ; hence, from (1.2) we find that  $E X = n/2$ , and also that  $\sigma^2 = \text{var}(X)$  is given by

$$\sigma^2 = \frac{n}{4}(1 - \lambda_{n,2,n}) + \frac{n^2}{4}\lambda_{n,2,n}. \tag{1.4}$$

To state our result for odd  $n$ , let

$$\bar{\lambda}_{n,b,r} = \begin{cases} \frac{1}{2}(\lambda_{n,b,m} + \lambda_{n,b,m+1}), & r \in \{m, m+1\}, \\ \lambda_{n,b,r}, & \text{otherwise,} \end{cases} \quad \text{and} \quad \bar{\lambda}_{n,b,n} = \prod_{r=1}^n \bar{\lambda}_{n,b,r},$$

that is,  $\bar{\lambda}_{n,b,n}$  is obtained from  $\lambda_{n,b,n}$  by replacing  $\lambda_{n,b,m}$  and  $\lambda_{n,b,m+1}$  in the product (1.1) by their average. In the odd case, we proceed by first coupling  $X$  to a more symmetric random variable  $V$  with mean and variance given respectively by

$$\mu_V = \frac{n}{2} \quad \text{and} \quad \sigma_V^2 = \frac{n}{4}(1 - \bar{\lambda}_{n,2,n}) + \frac{n^2}{4}\bar{\lambda}_{n,2,n}. \tag{1.5}$$

Then, with  $V$  in hand, we couple  $V$  to a variable with the  $V$ -size bias distribution, and proceed as in the even case. In Theorem 1.1, and the remainder of the paper,  $Z$  denotes a standard normal random variable.

**Theorem 1.1.** *Let  $X$  be the number of bulbs on at the terminal time  $n$  in the standard lightbulb process. Then, for all even  $n$ , with  $\sigma^2$  as given in (1.4),*

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P} \left( \frac{X - n/2}{\sigma} \leq z \right) - \mathbb{P}(Z \leq z) \right| \leq \frac{n}{2\sigma^2} \bar{\Delta}_0 + 1.64 \frac{n}{\sigma^3} + \frac{2}{\sigma} \quad \text{for all } n \geq 6, \quad (1.6)$$

where

$$\bar{\Delta}_0 = \frac{1}{2\sqrt{n}} + \frac{1}{2n} + \frac{1}{3} e^{-n/2},$$

and, for all odd  $n$ ,

$$\begin{aligned} \sup_{z \in \mathbb{R}} \left| \mathbb{P} \left( \frac{X - n/2}{\sigma_V} \leq z \right) - \mathbb{P}(Z \leq z) \right| \\ \leq \frac{n}{2\sigma_V^2} \bar{\Delta}_1 + 1.64 \frac{n}{\sigma_V^3} + \frac{2}{\sigma_V} \left( 1 + \frac{1}{\sqrt{2\pi}} \right) \quad \text{for all } n \geq 7, \end{aligned}$$

where  $\sigma_V^2$  is given in (1.5) and

$$\bar{\Delta}_1 = \frac{1}{\sqrt{n}} + \frac{1}{2\sqrt{2}} e^{-n/4}. \quad (1.7)$$

In the even case, as  $\lambda_{n,2,n}$  decays exponentially fast to 0, the variance  $\sigma^2$  is of order  $n$  and the bound (1.6), therefore, of order  $1/\sqrt{n}$ ; analogous remarks hold for the case where  $n$  is odd.

We now more formally describe the lightbulb process on  $n$  bulbs with  $k$  stages. With  $n \in \mathbb{N}$  fixed and  $s = (s_1, \dots, s_k)$  with  $s_r \in \{0, \dots, n\}$  for  $r = 1, \dots, k$ , we will let  $X_s = \{X_{rj} : r = 0, 1, \dots, k, j = 1, \dots, n\}$  denote a collection of Bernoulli variables. The initial state of the bulbs is given deterministically by  $\{X_{0j}, j = 1, \dots, n\}$ , which will be taken to be state zero, that is, all bulbs off, unless specifically stated otherwise; in fact, nonzero initial conditions are considered only in Corollary 4.1. For  $r \in \{1, \dots, k\}$ , the components of the switch variables  $X_s$  have the interpretation that

$$X_{rj} = \begin{cases} 1 & \text{if the status of bulb } j \text{ is changed at stage } r, \\ 0 & \text{otherwise.} \end{cases}$$

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  $e = \{e_{rj}\}_{1 \leq r \leq k, 1 \leq j \leq n}$  an array of  $\{0, 1\}$ -valued variables, the distribution of  $X_s$  is given by

$$\mathbb{P}(X_s = e) = \begin{cases} \prod_{r=1}^k \binom{n}{s_r}^{-1} & \text{if } \sum_{j=1}^n e_{rj} = s_r, r = 1, \dots, k, \\ 0 & \text{otherwise.} \end{cases} \quad (1.8)$$

Clearly, the vectors of stage  $r$  switch variables,  $(X_{r1}, \dots, X_{rn})$ , are exchangeable and the marginal distribution of the components  $X_{rj}$  are Bernoulli with success probability  $s_r/n$ . In general, for  $j = 1, \dots, n$ , the variables

$$X_j = \left( \sum_{r=0}^k X_{rj} \right) \bmod 2 \quad \text{and} \quad X_s = \sum_{j=1}^n X_j \quad (1.9)$$

are the indicator that bulb  $j$  is on at the terminal time  $k$ , and the total number of bulbs on at that time, respectively. For the standard lightbulb process, we will write  $X$  and  $X$  for  $X_n$  and  $X_n$ , respectively.

The lightbulb process, where the individual states of the  $n$  bulbs evolve according to the same marginal Markov chain, is a special case of a class of multivariate chains studied in [12], known as composition Markov chains of multinomial type. As shown in [12], such chains admit explicit full spectral decompositions, and in particular, the transition matrices for the stages of the lightbulb process can be simultaneously diagonalized by a Hadamard matrix. These properties were put to use in [8] for the calculation of the moments needed to compute the mean and variance of  $X$ . 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 derivation 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 [11]. By contrast, the method of moments requires the calculation and appropriate convergence of moments of all orders, and yields only convergence in distribution. Stein’s method for the normal is based on the characterization of the normal distribution in [10], which states that  $Z$  is a standard normal variable if and only if

$$E[Zg(Z)] = E[g'(Z)] \tag{1.10}$$

for all absolutely continuous functions  $g$  for which these expectations exist. The idea behind Stein’s method is that if a mean 0, variance 1 random variable  $W$  is close in distribution to  $Z$ , then  $W$  will satisfy (1.10) approximately. Hence, to gauge the proximity of  $W$  to  $Z$  for a given test function  $h$ , we 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  $E[f'(W) - Wf(W)]$ . A priori it may appear that an evaluation of  $E[f'(W) - Wf(W)]$  would be more difficult than that for  $Eh(W) - Nh$ . However, the former form may be handled through couplings.

Here we consider size bias couplings to evaluate  $E[f'(W) - Wf(W)]$ . Given a nonnegative random variable  $Y$  with positive finite mean  $\mu = EY$ , we say that  $Y^s$  has the  $Y$ -size bias distribution if  $P(Y^s \in dy) = (y/\mu)P(Y \in dy)$ , or more formally, if

$$E[Yg(Y)] = \mu E[g(Y^s)] \quad \text{for all bounded continuous functions } g. \tag{1.11}$$

The use of size bias couplings in Stein’s method was introduced in [1], where it was applied to derive 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 [6] 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 [4] 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. A more complete treatment of Stein’s method and its applications can be found in [2].

Here we prove and apply Theorem 2.1 below for bounded, monotone size bias couplings, which requires that the random variable  $Y$  of interest, and a random variable  $Y^s$ , having the  $Y$ -size bias distribution, be constructed on a common space such that, for some nonnegative constant  $B$ ,

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

with probability 1. 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  $E Y$ ,  $\text{var}(Y)$ ,  $B$ , and the quantity

$$\Delta = \sqrt{\text{var}(E[Y^s - Y \mid Y])}. \tag{1.12}$$

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

For the standard lightbulb process, a size bias coupling of  $X$  to  $X^s$  is achieved in the even case by the construction, for each  $i = 1, \dots, n$ , of a collection  $X^i$  from the given  $X$  as follows. Recalling (1.9), where, for  $s = n$ , we have  $k = n$ , if  $X_i = 1$ , that is, if bulb  $i$  is on at the terminal time, 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 with the values of  $X_{n/2,i}$  and  $X_{n/2,J}$  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  $1, \dots, n$ , the variable  $X^s = X^I$  has the  $X$ -size bias distribution, essentially due to the fact, shown in Lemma 3.2 below, that

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

Owing to the parity issue, to handle the odd case when  $n = 2m + 1$ , we first construct a coupling of  $X$  to a more symmetric variable  $V$ . The variable  $V$  is constructed by randomizing stages  $m$  and  $m + 1$  in the switch variables that yield  $X$ . In particular, at stage  $m$  we add an additional switch with probability  $\frac{1}{2}$  and, independently, at stage  $m + 1$  we remove an existing switch with probability  $\frac{1}{2}$ . A size bias coupling of  $V$  to  $V^s$  can be achieved as in the even case, thus yielding a bound to the normal for  $X$ . We remark that the size-biased couplings developed here are used in [3] to show that the distribution of  $X$ , in both the even and odd cases, obeys concentration-of-measure-type inequalities.

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 construction and the proof of the bound for the even case of the lightbulb process are given in Section 3.1. Symmetrization, that is, the construction of  $V$  from  $X$ , coupling constructions for  $V$ , and an outline of the proof of the bound in the odd case are given in Section 3.2. Calculations of the bounds on the variance  $\Delta$  in (1.12) require estimates on  $\lambda_{n,b,s}$  in (1.1). These estimates, given in Section 4, are based on the work of [12] and yield the spectral decomposition of the underlying transition matrices of the chain. Complete detailed calculations can be found in the technical report [7].

## 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 [9]. With  $Y$  having mean  $\mu$  and variance  $\sigma^2$ , both finite and positive, with some slight abuse of notation in the definition of  $W^s$ , we set

$$W = \frac{Y - \mu}{\sigma} \quad \text{and} \quad W^s = \frac{Y^s - \mu}{\sigma}. \tag{2.1}$$

**Lemma 2.1.** *Let  $Y$  be a nonnegative random variable with mean  $\mu$  and variance  $\sigma^2$ , both finite and positive, and let  $Y^s$  be given on the same space as  $Y$ , having the  $Y$ -size bias distribution, and satisfying  $Y^s \geq Y$  with probability 1. Then, with  $W$  and  $W^s$  given in (2.1), for any  $z \in \mathbb{R}$  and  $a > 0$ ,*

$$\frac{\mu}{\sigma} E[W^s - W] \mathbf{1}_{\{W^s - W \leq a\}} \mathbf{1}_{\{z \leq W \leq z+a\}} \leq a.$$

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

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

Then, using  $|f(w)| \leq a$  for all  $w \in \mathbb{R}$ ,  $\text{var}(W) = 1$ , and the Cauchy–Schwarz inequality to obtain the first inequality, followed by definition (2.1) and the size bias relation (1.11), we have

$$\begin{aligned} a &\geq E[Wf(W)] \\ &= \frac{1}{\sigma} E[Y - \mu] f\left(\frac{Y - \mu}{\sigma}\right) \\ &= \frac{\mu}{\sigma} E[f(W^s) - f(W)] \\ &= \frac{\mu}{\sigma} E \int_0^{W^s - W} f'(W + t) dt \\ &\geq \frac{\mu}{\sigma} E \int_0^{W^s - W} \mathbf{1}_{\{0 \leq t \leq a\}} \mathbf{1}_{\{z \leq W \leq z + a\}} f'(W + t) dt, \end{aligned}$$

where in the final inequality we have used  $W^s \geq W$  and  $f'(w) \geq 0$  for all  $w \in \mathbb{R}$ . Noting that  $f'(W + t) = \mathbf{1}_{\{z \leq W + t \leq z + 2a\}}$ , and that  $0 \leq t \leq a$  and  $z \leq W \leq z + a$  imply that  $z \leq W + t \leq z + 2a$ , we have

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

and, therefore, we obtain

$$\begin{aligned} a &\geq \frac{\mu}{\sigma} E \int_0^{W^s - W} \mathbf{1}_{\{0 \leq t \leq a\}} \mathbf{1}_{\{z \leq W \leq z + a\}} dt \\ &= \frac{\mu}{\sigma} E[\min(a, W^s - W) \mathbf{1}_{\{z \leq W \leq z + a\}}] \\ &\geq \frac{\mu}{\sigma} E[W^s - W] \mathbf{1}_{\{W^s - W \leq a\}} \mathbf{1}_{\{z \leq W \leq z + a\}}, \end{aligned}$$

as claimed.

**Theorem 2.1.** *Let  $Y$  be a nonnegative random variable with mean  $\mu$  and variance  $\sigma^2$ , both finite and positive, and let  $Y^s$  be given on the same space as  $Y$ , with the  $Y$ -size bias distribution, satisfying  $Y \leq Y^s \leq Y + B$  with probability 1, for some positive constant  $B$ . Then, with  $W$  and  $W^s$  given in (2.1), we have*

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

where

$$\Delta = \sqrt{\text{var}(E[Y^s - Y | Y])} \quad \text{and} \quad \delta = \frac{B}{\sigma}. \tag{2.2}$$

*Proof.* For  $z \in \mathbb{R}$  arbitrary, let  $h(w) = \mathbf{1}_{\{w \leq z\}}$  and let  $f(w)$  be the unique bounded solution to the Stein equation

$$f'(w) - wf(w) = h(w) - Nh, \tag{2.3}$$

where  $Nh = E h(Z)$ . Substituting  $W$  into (2.3), and using definition (2.1) and the size bias

relation (1.11) yields

$$\begin{aligned}
 & \mathbb{E}[h(W) - Nh] \\
 &= \mathbb{E}[f'(W) - Wf(W)] \\
 &= \mathbb{E}\left[f'(W) - \frac{\mu}{\sigma}(f(W^s) - f(W))\right] \\
 &= \mathbb{E}\left[f'(W)\left(1 - \frac{\mu}{\sigma}(W^s - W)\right) - \frac{\mu}{\sigma} \int_0^{W^s - W} (f'(W + t) - f'(W)) dt\right]. \tag{2.4}
 \end{aligned}$$

As compiled in Lemma 2.3 of [2], we have the following bounds on the solution  $f$  from Lemma 2 of [11, Chapter II]:

$$0 < f(w) < \frac{\sqrt{2\pi}}{4} \quad \text{and} \quad |f'(w)| \leq 1. \tag{2.5}$$

Also, as previously noted in [9], as a consequence of (2.5) and the mean value theorem, we obtain

$$|(w + t)f(w + t) - wf(w)| \leq \left(|w| + \frac{\sqrt{2\pi}}{4}\right)|t|. \tag{2.6}$$

Noting that  $\mathbb{E} Y^s = \mathbb{E} Y^2 / \mu$  by (1.11), we find that

$$\frac{\mu}{\sigma} \mathbb{E}[W^s - W] = \frac{\mu}{\sigma^2} \left(\frac{\mathbb{E} Y^2}{\mu} - \mu\right) = 1.$$

Therefore, taking the expectation by conditioning, and then applying (2.5) and the Cauchy–Schwarz inequality, we bound the first term in (2.4) as

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

To bound the remaining term of (2.4), using (2.3), we have

$$\begin{aligned}
 \frac{\mu}{\sigma} \int_0^{W^s - W} (f'(W + t) - f'(W)) dt &= \frac{\mu}{\sigma} \int_0^{W^s - W} [(W + t)f(W + t) - Wf(W)] dt \\
 &\quad + \frac{\mu}{\sigma} \int_0^{W^s - W} (\mathbf{1}_{\{W+t \leq z\}} - \mathbf{1}_{\{W \leq z\}}) dt. \tag{2.7}
 \end{aligned}$$

Applying (2.6) to the first term in (2.7), and using  $0 \leq W^s - W \leq \delta$  and  $\mathbb{E} W^2 = 1$ , shows that the absolute value of the expectation of this term is bounded by

$$\begin{aligned}
 \frac{\mu}{\sigma} \mathbb{E}\left[\int_0^{W^s - W} \left(|W| + \frac{\sqrt{2\pi}}{4}\right)t dt\right] &= \frac{\mu}{2\sigma} \mathbb{E}\left[(W^s - W)^2 \left(|W| + \frac{\sqrt{2\pi}}{4}\right)\right] \\
 &\leq \frac{\mu}{2\sigma} \delta^2 \left(1 + \frac{\sqrt{2\pi}}{4}\right) \\
 &\leq 0.82 \frac{\delta^2 \mu}{\sigma}.
 \end{aligned}$$

Taking the expectation of the absolute value of the second term in (2.7) we obtain

$$\begin{aligned} \frac{\mu}{\sigma} \mathbb{E} \left| \int_0^{W^s - W} (\mathbf{1}_{\{W+t \leq z\}} - \mathbf{1}_{\{W \leq z\}}) dt \right| &= \frac{\mu}{\sigma} \mathbb{E} \left[ \int_0^{W^s - W} \mathbf{1}_{\{z-t < W \leq z\}} dt \right] \\ &\leq \frac{\mu}{\sigma} \mathbb{E}[W^s - W] \mathbf{1}_{\{z-\delta < W \leq z\}}, \end{aligned}$$

again using the fact that  $0 \leq W^s - W \leq \delta$  with probability 1. Lemma 2.1 with  $a = \delta$  and  $z$  replaced by  $z - \delta$  shows that this term can be no more than  $\delta$ . Since  $z \in \mathbb{R}$  was arbitrary, the proof is complete.

### 3. Normal approximation of $X$

The next lemma shows that the size bias distribution of a sum may be achieved by taking certain mixtures. The result is a special case of Lemma 2.1 of [6], but we give a short direct proof to make the paper more self-contained.

**Lemma 3.1.** *Suppose that  $X$  is a sum of nontrivial exchangeable Bernoulli variables  $X_1, \dots, X_n$ , and that, for  $i \in \{1, \dots, n\}$ , the variables  $X_1^i, \dots, X_n^i$  have joint distribution*

$$\mathcal{L}(X_1^i, \dots, X_n^i) = \mathcal{L}(X_1, \dots, X_n \mid X_i = 1).$$

Then

$$X^i = \sum_{j=1}^n X_j^i$$

has the  $X$ -size bias distribution  $X^s$ , as does the mixture  $X^I$  when  $I$  is a random index with values in  $\{1, \dots, n\}$ , independent of all other variables.

*Proof.* For  $i \in \{1, \dots, n\}$ , we first need to show that  $X^i$  satisfies (1.11), that is, that  $\mathbb{E}[X] \mathbb{E}[g(X^i)] = \mathbb{E}[Xg(X)]$  holds for a given bounded continuous  $g$ . Now, for such  $g$ ,

$$\mathbb{E}[Xg(X)] = \sum_{j=1}^n \mathbb{E}[X_j g(X)] = \sum_{j=1}^n \mathbb{P}(X_j = 1) \mathbb{E}[g(X) \mid X_j = 1].$$

As exchangeability implies that  $\mathbb{E}[g(X) \mid X_j = 1] = \mathbb{E}[g(X) \mid X_i = 1]$  for all  $j = 1, \dots, n$ , we have

$$\mathbb{E}[Xg(X)] = \left( \sum_{j=1}^n \mathbb{P}(X_j = 1) \right) \mathbb{E}[g(X) \mid X_i = 1] = \mathbb{E}[X] \mathbb{E}[g(X^i)],$$

proving the first claim. The second claim now follows from

$$\begin{aligned} \mathbb{E} g(X^I) &= \sum_{i=1}^n \mathbb{E}[g(X^i), I = i] \\ &= \sum_{i=1}^n \mathbb{E}[g(X^i) \mid I = i] \mathbb{P}(I = i) \\ &= \sum_{i=1}^n \mathbb{E} g(X^i) \mathbb{P}(I = i) \end{aligned}$$

$$\begin{aligned} &= \sum_{i=1}^n \mathbb{E} g(X^s) \mathbb{P}(I = i) \\ &= \mathbb{E} g(X^s) \sum_{i=1}^n \mathbb{P}(I = i) \\ &= \mathbb{E} g(X^s). \end{aligned}$$

**3.1. Even case**

In this subsection we provide the proof of Theorem 1.1 for even  $n$ . We begin by describing a coupling of  $X$ , the total number of bulbs on at the terminal time  $n$  in the standard lightbulb process, to a variable  $X^s$  with the  $X$ -size bias distribution. Throughout, we let  $\mathcal{U}(S)$  denote the uniform distribution over a finite set  $S$ .

**Theorem 3.1.** *With even  $n \in \mathbb{N}$ , let the collection of switch variables  $\mathbf{X} = \{X_{rj} : r, j = 1, \dots, n\}$  and  $X$  satisfy (1.8) and (1.9), respectively, with  $s = n$ . For every  $i = 1, \dots, n$ , let  $\mathbf{X}^i$  be given from  $\mathbf{X}$  as follows. If  $X_i = 1$  then  $\mathbf{X}^i = \mathbf{X}$ . Otherwise, with  $\mathcal{L}(J^i | \mathbf{X}) = \mathcal{U}\{j : X_{n/2,j} = 1 - X_{n/2,i}\}$ , let  $\mathbf{X}^i = \{X_{rj}^i : r, j = 1, \dots, n\}$ , where*

$$X_{rj}^i = \begin{cases} X_{rj}, & r \neq n/2, \\ X_{n/2,j}, & r = n/2, j \notin \{i, J^i\}, \\ X_{n/2,J^i}, & r = n/2, j = i, \\ X_{n/2,i}, & r = n/2, j = J^i, \end{cases}$$

and let  $X^i = \sum_{j=1}^n X_j^i$ , where

$$X_j^i = \left( \sum_{r=1}^n X_{rj}^i \right) \bmod 2.$$

Then, with  $I$  uniformly chosen from  $\{1, \dots, n\}$  and independent of all other variables, the mixture  $X^I = X^s$  has the  $X$ -size bias distribution and satisfies

$$X^s - X = 2\mathbf{1}_{(X_I=0, X_{J^I}=0)} \quad \text{and} \quad X \leq X^s \leq X + 2. \tag{3.1}$$

To prove Theorem 3.1, we make use of a preliminary lemma, and also of the fact that

$$\mathbb{P}(X_j = 0) = \mathbb{P}(X_j = 1) = \frac{1}{2} \quad \text{for all } j = 1, \dots, n. \tag{3.2}$$

The equalities in (3.2) follow from  $\mathbb{E} X = n/2$ , itself implied by (1.2) and that  $\lambda_{1,n,n} = 0$ , as noted earlier.

**Lemma 3.2.** *For all  $i = 1, \dots, n$ , the collection of random variables  $\mathbf{X}^i$  constructed from  $\mathbf{X}$  as specified in Theorem 3.1 satisfies*

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

*Proof.* For a given  $i \in \{1, \dots, n\}$ , let  $\mathbf{e} = \{e_{rj} : r, j = 1, \dots, n\}$  with  $e_{rj} \in \{0, 1\}$  for  $r, j = 1, \dots, n$ . First note that, since  $\mathbf{X}^i = \mathbf{X}$  when  $X_i = 1$ , we have

$$\begin{aligned} \mathbb{P}(\mathbf{X}^i = \mathbf{e}) &= \mathbb{P}(X_i = 1) \mathbb{P}(\mathbf{X}^i = \mathbf{e} | X_i = 1) + \mathbb{P}(X_i = 0) \mathbb{P}(\mathbf{X}^i = \mathbf{e} | X_i = 0) \\ &= \mathbb{P}(X_i = 1) \mathbb{P}(\mathbf{X} = \mathbf{e} | X_i = 1) + \mathbb{P}(X_i = 0) \mathbb{P}(\mathbf{X}^i = \mathbf{e} | X_i = 0), \end{aligned}$$

so the desired conclusion is equivalent to

$$P(\mathbf{X}^i = \mathbf{e} \mid X_i = 0) = P(\mathbf{X} = \mathbf{e} \mid X_i = 1). \tag{3.3}$$

As the construction of  $\mathbf{X}^i$  preserves the number of switches in each stage  $r$ , we may assume that  $\sum_j e_{rj} = r$  for all  $r$ , as otherwise both sides of (3.3) are 0. If  $\sum_r e_{ri} = 0 \pmod 2$  then the left-hand side of (3.3) is 0 since  $X_i^i = 1$  by construction; similarly, the right-hand side is 0 as  $\mathbf{X} = \mathbf{e}$  implies that  $X_i = 0$ . Hence, we need only verify (3.3), assuming that

$$\sum_{j=1}^n e_{rj} = r \quad \text{for all } r = 1, \dots, n \quad \text{and} \quad \sum_{r=1}^n e_{ri} = 1 \pmod 2. \tag{3.4}$$

Writing  $J$  for  $J^i$  for simplicity, and letting  $\mathbf{e}^{i,j}$  denote the array  $\mathbf{e}$  with coordinates  $e_{n/2,i}$  and  $e_{n/2,j}$  interchanged, by (3.2) we have

$$\begin{aligned} P(\mathbf{X}^i = \mathbf{e} \mid X_i = 0) &= 2P(\mathbf{X}^i = \mathbf{e}, X_i = 0) \\ &= 2 \sum_{j=1}^n P(\mathbf{X}^i = \mathbf{e}, X_i = 0, J = j) \\ &= 2 \sum_{j=1}^n P(\mathbf{X} = \mathbf{e}^{i,j}, X_i = 0, J = j). \end{aligned}$$

Note that when  $e_{n/2,i} = e_{n/2,j}$ , or, equivalently,  $\mathbf{e}_{n/2,i}^{i,j} = \mathbf{e}_{n/2,j}^{i,j}$ , then

$$P(\mathbf{X} = \mathbf{e}^{i,j}, X_i = 0, J = j) = P(\mathbf{X} = \mathbf{e}^{i,j}, J = j), \tag{3.5}$$

as both sides are 0, since on  $J = j$  we have  $X_{n/2,i} \neq X_{n/2,j}$ . Otherwise,  $e_{n/2,i} \neq e_{n/2,j}$ , and, by the second equality in (3.4),

$$\sum_{r=1}^n \mathbf{e}_{ri}^{i,j} = \sum_{r \neq n/2} e_{ri} + \mathbf{e}_{n/2,i}^{i,j} = \sum_{r \neq n/2} e_{ri} + 1 - e_{n/2,i} = \sum_{r=1}^n e_{ri} + 1 = 0,$$

with equalities modulo 2, so (3.5) holds again. Hence,

$$\begin{aligned} P(\mathbf{X}^i = \mathbf{e} \mid X_i = 0) &= 2 \sum_{j=1}^n P(\mathbf{X} = \mathbf{e}^{i,j}, J = j) \\ &= 2 \sum_{j=1}^n P(J = j \mid \mathbf{X} = \mathbf{e}^{i,j}) P(\mathbf{X} = \mathbf{e}^{i,j}) \\ &= 2 \prod_{s=1}^n \binom{n}{s}^{-1} \sum_{j=1}^n P(J = j \mid \mathbf{X} = \mathbf{e}^{i,j}) \\ &= 2 \prod_{s=1}^n \binom{n}{s}^{-1}, \end{aligned}$$

where in the second to last equality we have used the fact that  $\mathbf{e}^{i,j}$ , as  $\mathbf{e}$ , satisfies the first equality of (3.4), and the distribution of  $\mathbf{X}$  given by (1.8), and in the last equality we have used the fact

that the sum of probabilities of the conditional distribution of  $J$  given an  $X$  configuration that satisfies the first equality of (3.4) for  $r = n/2$  must sum to 1. Now, again using the second equality in (3.4),

$$2 \prod_{s=1}^n \binom{n}{s}^{-1} = 2P(X = \mathbf{e}) = 2P(X = \mathbf{e}, X_i = 1) = P(X = \mathbf{e} \mid X_i = 1),$$

proving (3.3), and completing the proof of the lemma.

*Proof of Theorem 3.1.* That  $X^s$  has the  $X$ -size bias distribution follows from Lemmas 3.1 and 3.2. To prove the first equality in (3.1), note that if  $X_I = 1$  then  $X^I = X$ ; hence, in this case  $X^s = X$ . Otherwise,  $X_I = 0$  and the collection  $X^I$  is constructed from  $X$  by interchanging the stage  $n/2$ , unequal, switch variables  $X_{n/2,I}$  and  $X_{n/2,J^I}$ . If  $X_{J^I} = 1$  then after the interchange  $X_I^I = 1$  and  $X_{J^I}^I = 0$ , yielding  $X^s = X$ . If  $X_{J^I} = 0$  then after the interchange  $X_I^I = 1$  and  $X_{J^I}^I = 1$ , yielding  $X^s = X + 2$ . The second claim in (3.1) is an immediate consequence of the first.

The following lemma shows that, for the case at hand, the variance of the conditional expectation term (2.2) in Theorem 2.1 may be expressed in terms of quantities of the form

$$g_{\alpha,\beta} = P(X_1 = \dots = X_{\alpha+\beta} = 0, X_{n/2,1} = \dots = X_{n/2,\alpha} = 0, X_{n/2,\alpha+1} = \dots = X_{n/2,\alpha+\beta} = 1), \tag{3.6}$$

the probability that, when applying switch pattern  $\mathbf{n}$ , bulbs numbered 1 through  $\alpha + \beta$  terminate in the off position, and in stage  $n/2$ , bulbs numbered 1 through  $\alpha$  receive switch variable 0 and bulbs numbered  $\alpha + 1$  through  $\alpha + \beta$  receive switch variable 1.

Using the spectral decomposition in Section 4 to handle the probabilities in (3.6), we now provide an upper bound to term (2.2) when applying Theorem 2.1 for even  $n$ . Let

$$\mathbf{n}_{n/2} = \left(1, \dots, \frac{n}{2} - 1, \frac{n}{2} + 1, \dots, n\right), \tag{3.7}$$

the vector  $\mathbf{n}$  with its  $(n/2)$ th component deleted.

**Lemma 3.3.** *Let  $n$  be even, and let  $X$  and  $X^s$  be given by Theorem 3.1. Then, for  $n \geq 6$ ,*

$$\Delta_0 \leq \bar{\Delta}_0,$$

where

$$\Delta_0 = \sqrt{\text{var}(E[X^s - X \mid X])} \quad \text{and} \quad \bar{\Delta}_0 = \frac{1}{2\sqrt{n}} + \frac{1}{2n} + \frac{1}{3}e^{-n/2}.$$

*Proof.* We apply the construction of  $X^s$ , and the conclusions, of Theorem 3.1. For notational simplicity, let  $J^I = J$ , so in particular, from (3.1) we have  $X^s - X = 2\mathbf{1}_{\{X_I=0, X_J=0\}}$ . Expanding the indicator over the possible values of  $I$  and  $J$ , and then over the values of the

switch variables  $X_{n/2,i}$  and  $X_{n/2,j}$  yields

$$\begin{aligned} \mathbf{1}_{\{X_I=0, X_J=0\}} &= \sum_{i,j=1}^n \mathbf{1}_{\{X_i=0, X_j=0\}} \mathbf{1}_{\{I=i, J=j\}} \\ &= \sum_{i,j=1}^n \mathbf{1}_{\{X_i=0, X_j=0, X_{n/2,i}=0\}} \mathbf{1}_{\{I=i, J=j\}} \\ &\quad + \sum_{i,j=1}^n \mathbf{1}_{\{X_i=0, X_j=0, X_{n/2,i}=1\}} \mathbf{1}_{\{I=i, J=j\}} \\ &= \sum_{i \neq j} \mathbf{1}_{\{X_i=0, X_j=0, X_{n/2,i}=0, X_{n/2,j}=1\}} \mathbf{1}_{\{I=i, J=j\}} \\ &\quad + \sum_{i \neq j} \mathbf{1}_{\{X_i=0, X_j=0, X_{n/2,i}=1, X_{n/2,j}=0\}} \mathbf{1}_{\{I=i, J=j\}} \\ &= 2 \sum_{i \neq j} \mathbf{1}_{\{X_i=0, X_j=0, X_{n/2,i}=0, X_{n/2,j}=1\}} \mathbf{1}_{\{I=i, J=j\}}, \end{aligned}$$

where the second to last equality holds almost surely, as the probability of the event  $\{I = i, J = j\}$  is 0 whenever  $X_{n/2,i}$  and  $X_{n/2,j}$  agree, and the last inequality holds since the final expression is the sum of two terms which can be seen to be equal by reversing the roles of  $i$  and  $j$ .

To obtain a tractable bound on the required variance, we apply the inequality

$$\text{var}(\mathbb{E}[X^S - X \mid \mathcal{F}]) \leq \text{var}(\mathbb{E}[X^S - X \mid \mathcal{F}]), \tag{3.8}$$

which holds when  $\mathcal{F}$  is any  $\sigma$ -algebra with respect to which  $X$  is measurable (see [6] for example). Here we let  $\mathcal{F}$  be the  $\sigma$ -algebra generated by  $\mathbf{X}$ , the collection of all switch variables. The first indicator in the final sum above,  $\mathbf{1}_{\{X_i=0, X_j=0, X_{n/2,i}=0, X_{n/2,j}=1\}}$ , is measurable with respect to  $\mathcal{F}$ . For the second indicator, conditioning on  $\mathcal{F}$  yields

$$\mathbb{E}[\mathbf{1}_{\{I=i, J=j\}} \mid \mathcal{F}] = \mathbb{P}(I = i, J = j \mid \mathcal{F}) = \frac{2}{n^2} \mathbf{1}_{\{X_{n/2,i} \neq X_{n/2,j}\}},$$

as, for any  $i$ , chosen with probability  $1/n$ , there are  $n/2$  choices for  $j$  satisfying the condition in the indicator. Hence, recalling from (3.1) that  $X^S - X = 2\mathbf{1}_{\{X_I=0, X_J=0\}}$ , we have

$$\mathbb{E}[X^S - X \mid \mathcal{F}] = U_n, \quad \text{where} \quad U_n = \frac{4}{n^2} \sum_{i \neq j} \mathbf{1}_{\{X_i=0, X_j=0, X_{n/2,i}=0, X_{n/2,j}=1\}}, \tag{3.9}$$

and  $\Delta_0^2 \leq \text{var}(U_n)$  by (3.8).

Taking the expectation of  $U_n$  in (3.9), using the exchangeability of the  $(n)_2$  terms in the sum and applying Corollary 4.1, also recalling the notation defined in (3.7), we have

$$\mathbb{E} U_n = \frac{4}{n^2} (n)_2 g_{1,1} = \frac{1}{4} (1 - \lambda_{n,2,n/2}). \tag{3.10}$$

Squaring (3.9) 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  $\{i_1, i_2\} \cap \{j_1, j_2\} = \emptyset$ , so  $|\{i_1, i_2, j_1, j_2\}| \in \{2, 3, 4\}$ , and we may write

$$U_n^2 = U_{n,2}^2 + U_{n,3}^2 + U_{n,4}^2, \tag{3.11}$$

where

$$U_{n,p}^2 = \frac{16}{n^4} \sum_{\substack{\{|i_1, i_2, j_1, j_2\}|=p \\ \{i_1, i_2\} \cap \{j_1, j_2\} = \emptyset}} \mathbf{1}_{\{X_{i_1}=0, X_{j_1}=0, X_{n/2, i_1}=0, X_{n/2, j_1}=1\}} \mathbf{1}_{\{X_{i_2}=0, X_{j_2}=0, X_{n/2, i_2}=0, X_{n/2, j_2}=1\}}.$$

Beginning the calculation with the main term  $U_{n,4}^2$ , where all four indices are distinct, taking the expectation using exchangeability, and applying Corollary 4.1 yields

$$E U_{n,4}^2 = \frac{16}{n^4} (n)_{4g_{2,2}} = \left(\frac{n-2}{4n}\right)^2 (1 - 2\lambda_{n,2,n_{n/2}} + \lambda_{n,4,n_{n/2}}). \tag{3.12}$$

With the inequalities over the summation in (3.11) in force, the event  $|\{i_1, i_2, j_1, j_2\}| = 3$  can only occur when

- (a)  $i_1 \neq i_2$  and  $j_1 = j_2$ , or
- (b)  $i_1 = i_2$  and  $j_1 \neq j_2$ .

Applying Corollary 4.1, case (a) leads to a contribution of

$$\frac{16}{n^4} (n)_{3g_{2,1}} = \frac{n-2}{4n^2} (1 + \lambda_{n,1,n_{n/2}} - \lambda_{n,2,n_{n/2}} - \lambda_{n,3,n_{n/2}}),$$

while in the same manner, again using Corollary 4.1, the contribution from case (b) is

$$\frac{16}{n^4} (n)_{3g_{1,2}} = \frac{n-2}{4n^2} (1 - \lambda_{n,1,n_{n/2}} - \lambda_{n,2,n_{n/2}} + \lambda_{n,3,n_{n/2}}).$$

Totalling we find that

$$E U_{n,3}^2 = \frac{n-2}{2n^2} (1 - \lambda_{n,2,n_{n/2}}). \tag{3.13}$$

With the inequalities over the summation in (3.11) 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 by Corollary 4.1,

$$E U_{n,2}^2 = \frac{16}{n^4} (n)_{2g_{1,1}} = \frac{1}{n^2} (1 - \lambda_{n,2,n_{n/2}}). \tag{3.14}$$

Summing (3.12), (3.13), and (3.14) we obtain

$$E U_n^2 = \left(\frac{n-2}{4n}\right)^2 (1 - 2\lambda_{n,2,n_{n/2}} + \lambda_{n,4,n_{n/2}}) + \frac{1}{2n} (1 - \lambda_{n,2,n_{n/2}}).$$

Subtracting the square of the first moment, given in (3.10), yields

$$\begin{aligned} \text{var}(U_n) &= \frac{1}{16} \left(1 - \frac{2}{n}\right)^2 (1 - 2\lambda_{n,2,n_{n/2}} + \lambda_{n,4,n_{n/2}}) + \frac{1}{2n} (1 - \lambda_{n,2,n_{n/2}}) \\ &\quad - \frac{1}{16} (1 - \lambda_{n,2,n_{n/2}})^2 \\ &= \frac{1}{16} (\lambda_{n,4,n_{n/2}} - \lambda_{n,2,n_{n/2}}^2) + \frac{1-n}{4n^2} (1 - 2\lambda_{n,2,n_{n/2}} + \lambda_{n,4,n_{n/2}}) \\ &\quad + \frac{1}{2n} (1 - \lambda_{n,2,n_{n/2}}) \\ &= \frac{1}{4n} + \frac{1}{4n^2} + \frac{1}{16} (\lambda_{n,4,n_{n/2}} - \lambda_{n,2,n_{n/2}}^2) - \frac{1}{2n^2} \lambda_{n,2,n_{n/2}} + \frac{1-n}{4n^2} \lambda_{n,4,n_{n/2}}. \end{aligned}$$

Now applying Lemma 4.5 given in Section 4, for  $n \geq 6$ , we obtain

$$\begin{aligned} \text{var}(U_n) &\leq \frac{1}{4n} + \frac{1}{4n^2} + \frac{1}{16} \left( \frac{1}{2}e^{-n} + e^{-2n} \right) + \frac{1}{2n^2}e^{-n} + \frac{n-1}{8n^2}e^{-n} \\ &\leq \frac{1}{4n} + \frac{1}{4n^2} + e^{-n} \left( \frac{1}{16} + \frac{1}{n^2} + \frac{1}{8n} \right). \end{aligned}$$

The inequality  $\sqrt{a+b+c} \leq \sqrt{a} + \sqrt{b} + \sqrt{c}$ , holding for all nonnegative  $a, b$ , and  $c$ , now yields the claim of the lemma.

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  $B = 2$ , by the second inequality in (3.1). Hence, the result for the even case follows by applying Theorem 2.1 with  $\mu = n/2$ ,  $\delta = 2/\sigma$ , and the bound  $\Delta_0$  on  $\Delta_0$  given in Lemma 3.3.

### 3.2. Odd case

Now we move to the case where  $n = 2m + 1$  is odd. Instead of directly forming a size-biased coupling 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  $m + 1$ . In particular, at stage  $m$  we add an additional switch with probability  $\frac{1}{2}$  and, independently at stage  $m + 1$  we remove an existing switch with probability  $\frac{1}{2}$ .

Formally, let  $X = \{X_{rj} : r, j = 1, \dots, n\}$  be a collection of switch variables with distribution given by (1.8) with  $s = n$ , and let  $X = X_n$  be given by (1.9) with  $k = n$ . Let

$$\mathcal{L}(B_m \mid X) = \mathcal{U}\{j : X_{mj} = 0\} \quad \text{and} \quad \mathcal{L}(B_{m+1} \mid X) = \mathcal{U}\{j : X_{m+1,j} = 1\}$$

with  $B_m$  and  $B_{m+1}$  conditionally independent given  $X$ , and let  $C_m$  and  $C_{m+1}$  be symmetric Bernoulli variables, independent of  $X$  and of  $B_m$  and  $B_{m+1}$ . Now let a collection of switch variables  $V = \{V_{rj}, r, j = 1, \dots, n\}$  be defined by

$$V_{rj} = \begin{cases} X_{rj}, & r \notin \{m, m + 1\}, \\ X_{mj}, & r = m, j \neq B_m, \\ C_m, & r = m, j = B_m, \\ X_{m+1,j}, & r = m + 1, j \neq B_{m+1}, \\ C_{m+1}, & r = m + 1, j = B_{m+1}, \end{cases} \tag{3.15}$$

and set

$$V = \sum_{j=1}^n V_j, \quad \text{where} \quad V_j = \left( \sum_{r=1}^n V_{rj} \right) \text{ mod } 2. \tag{3.16}$$

In other words, in all stages other than  $m$  and  $m + 1$  the switch variables that produce  $V$  are those from the given collection  $X$ . In stage  $m$ , the switch variables for all bulbs but bulb  $B_m$ , chosen uniformly over all bulbs in that stage that were not toggled, are those given by  $X$ . The switch variable for  $B_m$  in stage  $m$ , however, is set to  $C_m$ , which takes the values 0 and 1 equally likely. Hence, with probability  $\frac{1}{2}$ , one additional bulb in stage  $m$  is toggled. Similarly, in stage  $m + 1$ , the switch variable of bulb  $B_{m+1}$ , uniformly selected from all the bulbs that were toggled

in that stage, is no longer toggled with probability  $\frac{1}{2}$ . Since  $X$  and  $V$  differ in at most two switches, we have

$$|X - V| \leq 2. \tag{3.17}$$

It is now not difficult to show that the collections of variables  $\{V_{rj}, j = 1, \dots, n\}$  are mutually independent for  $r = 1, \dots, n$ , and that the distribution of the switch variables  $V$  in stages  $m$  and  $m + 1$  are the same equal mixture, that is, for  $r \in \{m, m + 1\}$ ,

$$\mathcal{L}(V_{r1}, \dots, V_{rn}) = \frac{1}{2}\mathcal{L}(X_{m1}, \dots, X_{mn}) + \frac{1}{2}\mathcal{L}(X_{m+1,1}, \dots, X_{m+1,n}).$$

The expressions in (1.5) for the mean and variance of  $V$  now follow from (1.2); see [7] for details.

We now present a size bias coupling for  $V$ . As in the even case, the variable  $V^s$  is obtained by first constructing, for each  $i = 1, \dots, n$ , switch variables  $V^i$  that satisfy

$$\mathcal{L}(V^i) = \mathcal{L}(V \mid V_i = 1). \tag{3.18}$$

For a given  $i = 1, \dots, n$ , to construct  $V^i$ , we first determine if  $V_i = 1$ . If so, set  $V^i = V$ . Otherwise, let  $M$  be a variable that chooses from the stages  $m$  and  $m + 1$  uniformly and independently of  $V$ . Since in this case  $V_i = 0$ , we may achieve  $V_i^i = 1$  by changing the switch variable  $V_{Mi}$  to  $1 - V_{Mi}$ . The coupling accomplishes this change in one of two possible ways.

To introduce the first way, called a flip, we say that a configuration  $e$  of binary switch variable values is feasible if  $P(V = e) \neq 0$ , that is, when

$$\sum_{j=1}^n e_{rj} = r \quad \text{for } r \notin \{m, m + 1\} \quad \text{and} \quad \sum_{j=1}^n e_{rj} \in \{m, m + 1\} \quad \text{for } r \in \{m, m + 1\}.$$

If flipping  $V_{Mi}$  to  $1 - V_{Mi}$  results in a feasible configuration, then the flip is made with probability  $1/(m + 1)$ . In other words, given  $e, r \in \{m, m + 1\}$  and  $i = 1, \dots, n$ , let  $e^{r,i}$  be the configuration with entries

$$e_{sl}^{r,i} = \begin{cases} e_{sl}, & s \neq r \text{ or } l \neq i, \\ 1 - e_{ri}, & s = r \text{ and } l = i. \end{cases}$$

Defining  $V^{M,i}$  in a like manner, the distribution of  $F^i$ , the indicator that  $V_{Mi}$  is flipped, is given by

$$P(F^i = 1 \mid V, M) = \frac{1}{m + 1} \mathbf{1}_{\{V^{M,i} \text{ is feasible}\}}. \tag{3.19}$$

If the flip is unsuccessful, that is, if  $F^i = 0$ , we perform an ‘interchange’ in stage  $M$ , much like the coupling in the even case. For a configuration  $e, r \in \{m, m + 1\}$  and  $i, j \in \{1, \dots, n\}$ , let

$$e_{sl}^{r,i,j} = \begin{cases} e_{sl}, & s \neq r \text{ or } l \neq i, \\ e_{rl}, & s = r, l \notin \{i, j\}, \\ e_{ri}, & s = r, l = j, \\ e_{rj}, & s = r, l = i, \end{cases}$$

that is,  $e^{r,i,j}$  is the configuration  $e$  with the variables in the  $r, i$  and  $r, j$  positions interchanged. Now let  $J^i$  be a random index with distribution given by

$$\mathcal{L}(J^i \mid V, M) = \mathcal{U}\{j : V_{Mj} \neq V_{Mi}\}. \tag{3.20}$$

Defining  $V^{M,i,j}$  in a like manner, when  $F^i = 0$ , we interchange  $V_{Mi}$  with  $V_{M,J^i}$ . Hence, overall the configuration  $V^i$  is specified by

$$V^i = \begin{cases} V, & V_i = 1, \\ V^{M,i}, & V_i = 0, F^i = 1, \\ V^{M,i,J^i}, & V_i = 0, F^i = 0. \end{cases} \tag{3.21}$$

The following theorem shows that  $V^i$  satisfies (3.18). In the following we denote  $F^i$  and  $J^i$  by  $F$  and  $J$ , respectively, for simplicity.

**Theorem 3.2.** *Let  $V$  be constructed from  $X$  as in (3.15), let  $M$  be a random variable uniformly distributed over  $\{m, m + 1\}$  independent of  $V$ , and, for  $i \in \{1, \dots, n\}$ , let  $F$  and  $J$  have distributions as specified in (3.19) and (3.20), respectively. Then  $V^i$  given by (3.21) satisfies (3.18).*

Furthermore, letting  $V^i = \sum_{j=1}^n V_j^i$ , where

$$V_j^i = \left( \sum_{r=1}^n V_{rj}^i \right) \text{ mod } 2,$$

and  $I$  be uniformly chosen from  $\{1, \dots, n\}$  and independent of all other variables, the mixture  $V^I = V^s$  has the  $V$ -size bias distribution and satisfies

$$V^s - V = \mathbf{1}_{\{V_I=0, F=1\}} + 2\mathbf{1}_{\{V_I=0, X_J=0, F=0\}} \quad \text{and} \quad V \leq V^s \leq V + 2.$$

For a proof of Theorem 3.2, see [7]. The coupling so constructed leads to the following bound to the normal for  $V$ .

**Theorem 3.3.** *If  $n$  is odd and  $V$  is given by (3.16), then*

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P} \left( \frac{V - n/2}{\sigma_V} \leq z \right) - \mathbb{P}(Z \leq z) \right| \leq \frac{n}{2\sigma_V^2} \bar{\Delta}_1 + 1.64 \frac{n}{\sigma_V^3} + \frac{2}{\sigma_V} \quad \text{for all } n \geq 7,$$

where  $\sigma_V^2$  and  $\bar{\Delta}_1$  are given in (1.5) and (1.7), respectively.

*Outline of the proof.* By Theorem 3.2, we may apply Theorem 2.1 with  $\delta = 2/\sigma_V$ , and it only remains to prove that  $\bar{\Delta}_1$  in (1.7) is an upper bound on  $\Delta_1 = \sqrt{\text{var } E[V^s - V | V]}$  for all  $n \geq 7$ . The desired bound may be obtained by computations similar to, but somewhat more involved than, those for the even case; see [7] for details.

We now provide a bound for the normal approximation of  $X$  in the odd case. We remark that, using  $V$ , fewer error terms, and, therefore, a smaller bound, result 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  $W = (X - n/2)/\sigma_V$  and  $W_V = (V - n/2)/\sigma_V$ , recalling that  $|X - V| \leq 2$  from (3.17), we have

$$|W - W_V| = \frac{|X - V|}{\sigma_V} \leq \frac{2}{\sigma_V}. \tag{3.22}$$

With  $\Phi(z) = P(Z \leq z)$  and

$$C_V = \frac{n}{2\sigma_V^2} \bar{\Delta}_1 + 1.64 \frac{n}{\sigma_V^3} + \frac{2}{\sigma_V},$$

by (3.22) and Theorem 3.3, we obtain

$$\begin{aligned} P(W \leq z) - \Phi(z) &\leq P\left(W_V - \frac{2}{\sigma_V} \leq z\right) - \Phi(z) \\ &= P\left(W_V \leq z + \frac{2}{\sigma_V}\right) - \Phi\left(z + \frac{2}{\sigma_V}\right) + \Phi\left(z + \frac{2}{\sigma_V}\right) - \Phi(z) \\ &\leq C_V + \frac{2}{\sigma_V \sqrt{2\pi}}. \end{aligned}$$

As a corresponding lower bound can be similarly demonstrated, the claim is shown.

### 4. Spectral decomposition

In [12] 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  $\mathbf{P}$  that 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  $\mathbf{P}$ .

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

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

where we let  $\mathbf{e}_0 = (1, 0)^\top$  and  $\mathbf{e}_1 = (0, 1)^\top$  denote the 0 and 1 states of the bulb, for off and on, respectively. With  $b \in \{0, 1, \dots, n\}$ , let  $\mathbf{P}_{n,b,s}$  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. Let  $\mathbf{P}_{n,0,s} = \mathbf{I}$  for all  $n$  and  $s$ , and let  $\mathbf{I}_2$  be the  $2 \times 2$  identity matrix. For  $n \geq 1$ , the matrix  $\mathbf{P}_{n,b,s}$  is given recursively by

$$\mathbf{P}_{n,b,s} = \frac{s}{n} (\mathbf{P} \otimes \mathbf{P}_{n-1,b-1,s-1}) + \left(1 - \frac{s}{n}\right) (\mathbf{I}_2 \otimes \mathbf{P}_{n-1,b-1,s}) \quad \text{for } b \in \{1, \dots, n\},$$

as any particular bulb among the  $b$  in the subset considered is selected with probability  $s/n$  to undergo transition according to  $\mathbf{P}$ , 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  $\mathbf{P}$  is easily diagonalizable by the orthogonal matrix  $\mathbf{T}$  as

$$\mathbf{P} = \mathbf{T}^\top \mathbf{\Gamma} \mathbf{T}, \quad \text{where } \mathbf{T} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \quad \text{and } \mathbf{\Gamma} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}; \tag{4.1}$$

hence,  $\mathbf{P}_{n,b,s}$  is diagonalized by

$$\mathbf{P}_{n,b,s} = \bigotimes_b \mathbf{T}^\top \mathbf{\Gamma}_{n,b,s} \bigotimes_b \mathbf{T}, \tag{4.2}$$

where  $\Gamma_{n,0,s} = 1$  and  $\Gamma_{n,b,s}$  is given recursively by

$$\Gamma_{n,b,s} = \frac{s}{n}(\Gamma \otimes \Gamma_{n-1,b-1,s-1}) + \left(1 - \frac{s}{n}\right)(I_2 \otimes \Gamma_{n-1,b-1,s}) \quad \text{for } b \in \{1, \dots, n\}. \quad (4.3)$$

The next result describes the diagonal matrices  $\Gamma_{n,b,s}$  more explicitly in terms of a sequence of vectors  $\mathbf{a}_b$  of length  $2^b$  for all  $b \geq 1$  defined through the recursion

$$\mathbf{a}_b = (\mathbf{a}_{b-1}, \mathbf{a}_{b-1} + \mathbf{1}_{b-1}) \quad \text{for } b \geq 2, \text{ with } \mathbf{a}_1 = (0, 1), \quad (4.4)$$

where  $\mathbf{1}_b = (1, \dots, 1)$  is of size  $2^b$ . For example,

$$\mathbf{a}_1 = (0, 1), \quad \mathbf{a}_2 = (0, 1, 1, 2), \quad \text{and} \quad \mathbf{a}_3 = (0, 1, 1, 2, 1, 2, 2, 3). \quad (4.5)$$

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

**Lemma 4.1.** *For  $n \in \{0, 1, \dots, \}$ ,  $b, s \in \{0, \dots, n\}$ , and  $\lambda_{n,b,s}$  given by (1.1), the matrix  $\Gamma_{n,b,s}$  in (4.2) satisfies*

$$\Gamma_{n,b,s} = \text{diag}(\lambda_{n,a_1,s}, \dots, \lambda_{n,a_{2^b},s}).$$

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

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

For instance, from (4.5), for  $b = 2$ , we have

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

and, for  $b = 3$ ,

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

*Proof of Lemma 4.1.* As  $a_1 = 0$ , we have  $\Gamma_{n,0,s} = 1 = \lambda_{n,0,s}$ , so the lemma is true for  $b = 0$ . For the inductive step, assuming that the lemma is true for  $b - 1$ , by (4.3) and the definition of  $\Gamma$  from (4.1), it suffices to verify that

$$\frac{s}{n}\lambda_{n-1,a,s-1} + \left(1 - \frac{s}{n}\right)\lambda_{n-1,a,s} = \lambda_{n,a,s}$$

and

$$-\frac{s}{n}\lambda_{n-1,a,s-1} + \left(1 - \frac{s}{n}\right)\lambda_{n-1,a,s} = \lambda_{n,a+1,s}$$

for all  $a = 0, 1, \dots$ . To prove the first equality, by (1.1) we have

$$\begin{aligned} & \frac{s}{n}\lambda_{n-1,a,s-1} + \left(1 - \frac{s}{n}\right)\lambda_{n-1,a,s} \\ &= \frac{s}{n} \sum_{t=0}^a \binom{a}{t} (-2)^t \frac{(s-1)_t}{(n-1)_t} + \left(1 - \frac{s}{n}\right) \sum_{t=0}^a \binom{a}{t} (-2)^t \frac{(s)_t}{(n-1)_t} \\ &= \sum_{t=0}^a \binom{a}{t} (-2)^t \left( \frac{s}{n} \frac{(s-1)_t}{(n-1)_t} + \left(1 - \frac{s}{n}\right) \frac{(s)_t}{(n-1)_t} \right) \\ &= \sum_{t=0}^a \binom{a}{t} (-2)^t \frac{(s)_{t+1} + (n-s)(s)_t}{(n)_{t+1}} \end{aligned}$$

$$\begin{aligned}
 &= \sum_{t=0}^a \binom{a}{t} (-2)^t \frac{\binom{s}{t} (s-t+n-s)}{\binom{n}{t+1}} \\
 &= \sum_{t=0}^a \binom{a}{t} (-2)^t \frac{\binom{s}{t} (n-t)}{\binom{n}{t+1}} \\
 &= \sum_{t=0}^a \binom{a}{t} (-2)^t \frac{\binom{s}{t}}{\binom{n}{t}} \\
 &= \lambda_{n,a,s}.
 \end{aligned}$$

The second equality can be shown in a similar, though slightly more involved, fashion.

We note that [12] expresses these eigenvalues in terms of the hypergeometric function.

If the  $k$  stages of the process  $1, \dots, k$  use switches  $s = (s_1, \dots, s_k)$ , then since the matrices  $P_{n,b,s}$ ,  $s \in \{0, 1, \dots, n\}$ , are simultaneously diagonalizable by (4.2), the transition matrix  $P_{n,b,s}$  for any subset of  $b$  bulbs can be diagonalized as

$$P_{n,b,s} = \prod_{j=1}^k P_{n,b,s_j} = \bigotimes_{j=1}^b T^\top \Gamma_{n,b,s} \bigotimes_{j=1}^b T = \bigotimes_{j=1}^b T^\top \text{diag}(\lambda_{n,a_1,s}, \dots, \lambda_{n,a_b,s}) \bigotimes_{j=1}^b T, \tag{4.6}$$

where  $\lambda_{n,a,s}$  is given in (1.1) and

$$\Gamma_{n,b,s} = \prod_{j=1}^k \Gamma_{n,b,s_j}. \tag{4.7}$$

If  $\pi$  is a permutation of  $\{1, \dots, k\}$ , let  $\pi(s) = (s_{\pi(1)}, \dots, s_{\pi(k)})$ . As all the matrices  $\Gamma_{n,b,s}$  are diagonal, from (4.7) we have  $\Gamma_{n,b,s} = \Gamma_{n,b,\pi(s)}$ , and now from (4.6) we have the following result.

**Lemma 4.2.** *The distribution of the lightbulb chain is independent of the order in which the switch variables  $s$  are applied, that is, for all permutations  $\pi$ ,*

$$P_{n,b,s} = P_{n,b,\pi(s)}.$$

The following lemma helps us compute probabilities such as  $g_{\alpha,\beta,s}^{(l)}$  in (3.6). For  $j \in \{0, 1, \dots\}$ , let  $\Omega_{b,j}$  be the  $2^b \times 2^b$  diagonal matrix in the variables  $x_k$ ,  $k \in \{0, 1, \dots\}$ , given by

$$\Omega_{b,j} = \text{diag}(x_{a_1+j}, \dots, x_{a_b+j}), \tag{4.8}$$

and set

$$\mathbf{u}_b = \bigotimes_{j=1}^b T \mathbf{e}_0^{\otimes b} \quad \text{and} \quad \mathbf{w}_b = \bigotimes_{j=1}^b T \mathbf{e}_1^{\otimes b}, \tag{4.9}$$

where we recall that  $\mathbf{e}_0 = (1, 0)^\top$  and  $\mathbf{e}_1 = (0, 1)^\top$ . Note that, for  $b = 1$ , we have

$$\mathbf{u}_1 = T \mathbf{e}_0 = \frac{1}{\sqrt{2}}(1, -1)^\top \quad \text{and} \quad \mathbf{w}_1 = T \mathbf{e}_1 = \frac{1}{\sqrt{2}}(1, 1)^\top. \tag{4.10}$$

**Lemma 4.3.** *Let  $t \in \{0, 1, \dots\}$  and  $\Omega_t = \Omega_{t,0}$ , and suppose that, for some vector  $\mathbf{v}_t \in \mathbb{R}^{2^t}$ ,*

$$\mathbf{v}_t^\top \Omega_t \mathbf{u}_t = \frac{1}{2^t} \sum_{j=0}^t \binom{t}{j} a(j) x_j \tag{4.11}$$

*holds for  $t = b - 1$  with some sequence  $a(j)$ ,  $j = 0, \dots, b - 1$ . Then (4.11) holds for  $t = b$  when replacing  $\mathbf{v}_t$  by  $\mathbf{v}_b = \mathbf{u}_1 \otimes \mathbf{v}_t$  and  $a(j)$  by*

$$a_u(j) = \frac{b-j}{b} a(j) + \frac{j}{b} a(j-1), \tag{4.12}$$

*and for  $t = b$  when replacing  $\mathbf{v}_t$  by  $\mathbf{v}_b = \mathbf{w}_1 \otimes \mathbf{v}_t$  and  $a(j)$  by*

$$a_w(j) = \frac{b-j}{b} a(j) - \frac{j}{b} a(j-1). \tag{4.13}$$

*Proof.* By (4.4) we may write

$$\Omega_b = \begin{bmatrix} \Omega_{b-1,0} & \mathbf{0} \\ \mathbf{0} & \Omega_{b-1,1} \end{bmatrix}$$

and by (4.10) we have

$$\mathbf{u}_b = \mathbf{u}_1 \otimes \mathbf{u}_{b-1} = \frac{1}{\sqrt{2}} (\mathbf{u}_{b-1}^\top, -\mathbf{u}_{b-1}^\top)^\top.$$

Hence, when  $\mathbf{v}_b = \mathbf{u}_1 \otimes \mathbf{v}_{b-1} = (\mathbf{v}_{b-1}^\top, -\mathbf{v}_{b-1}^\top)^\top / \sqrt{2}$ , we obtain

$$\begin{aligned} \mathbf{v}_b^\top \Omega_b \mathbf{u}_b &= \frac{1}{2} (\mathbf{v}_{b-1}^\top \Omega_{b-1,0} \mathbf{u}_{b-1} + \mathbf{v}_{b-1}^\top \Omega_{b-1,1} \mathbf{u}_{b-1}) \\ &= \frac{1}{2^b} \sum_{j=0}^{b-1} \binom{b-1}{j} a(j) x_j + \frac{1}{2^b} \sum_{j=0}^{b-1} \binom{b-1}{j} a(j) x_{j+1} \\ &= \frac{1}{2^b} \sum_{j=0}^{b-1} \binom{b-1}{j} a(j) x_j + \frac{1}{2^b} \sum_{j=1}^b \binom{b-1}{j-1} a(j-1) x_j \\ &= \frac{1}{2^b} \sum_{j=0}^b \left( \binom{b-1}{j} a(j) + \binom{b-1}{j-1} a(j-1) \right) x_j \\ &= \frac{1}{2^b} \sum_{j=0}^b \binom{b}{j} \left( \left( \frac{b-j}{b} \right) a(j) + \left( \frac{j}{b} \right) a(j-1) \right) x_j \\ &= \frac{1}{2^b} \sum_{j=0}^b \binom{b}{j} a_u(j) x_j, \end{aligned}$$

as claimed. The proof is essentially the same, using (4.10), when  $\mathbf{v}_b = \mathbf{w}_1 \otimes \mathbf{v}_{b-1} = (\mathbf{v}_{b-1}^\top, \mathbf{v}_{b-1}^\top)^\top / \sqrt{2}$ .

In Corollary 4.1 below we use the conditional probability

$$\begin{aligned} f_{\alpha,\beta,s} &= P(X_i = 0, i = 1, \dots, \alpha + \beta \mid X_{0,i} = 0, i = 1, \dots, \alpha, \\ &\quad X_{0,i} = 1, i = \alpha + 1, \dots, \alpha + \beta) \end{aligned} \tag{4.14}$$

(see Lemma 4.4 below), where  $X_s$  has distribution given by (1.8), to express the functions  $g_{\alpha,\beta}$  in (3.6).

**Lemma 4.4.** For given  $\alpha, \beta \geq 0$ , setting  $b = \alpha + \beta$ , the probability  $f_{\alpha,\beta,s}$  in (4.14) is given by

$$f_{\alpha,\beta,s} = \frac{1}{2^b} \sum_{j=0}^b \binom{b}{j} a_{\alpha,\beta}(j) \lambda_{n,j,s}, \tag{4.15}$$

where  $a_{\alpha,0}(j) = 1$  for all  $\alpha \geq 0$  and

$$a_{\alpha,\beta}(j) = \frac{b-j}{b} a_{\alpha,\beta-1}(j) - \frac{j}{b} a_{\alpha,\beta-1}(j-1) \text{ for all } \alpha \geq 0, \beta \geq 1. \tag{4.16}$$

*Proof.* Using exchangeability for the first equality, extracting the relevant component of the  $k$ -step transition matrix and applying (4.6), we obtain

$$\begin{aligned} f_{\alpha,\beta,s} &= ((e_1^\top)^{\otimes \beta} \otimes (e_0^\top)^{\otimes \alpha}) \mathbf{P}_{n,b,s} e_0^{\otimes b} \\ &= ((e_1^\top)^{\otimes \beta} \otimes (e_0^\top)^{\otimes \alpha}) \bigotimes_b \mathbf{T}^\top \Gamma_{n,b,s} \bigotimes_b \mathbf{T} e_0^{\otimes b} \\ &= \mathbf{v}_b^\top \Gamma_{n,b,s} \mathbf{u}_b, \end{aligned}$$

where  $\mathbf{u}_\alpha$  and  $\mathbf{w}_\beta$  are given as in (4.9) and  $\mathbf{v}_b = \mathbf{w}_\beta \otimes \mathbf{u}_\alpha$ . Hence, with  $\Omega_b = \Omega_{b,0}$  as in (4.8), the result follows from

$$\mathbf{v}_b^\top \Omega_b \mathbf{u}_b = \frac{1}{2^b} \sum_{j=0}^b \binom{b}{j} a_{\alpha,\beta}(j) x_j. \tag{4.17}$$

We first prove the case in which  $\beta = 0$  by induction in  $\alpha$ ; note that in this case  $\mathbf{v}_b = \mathbf{u}_b$ . Equality (4.17) holds with  $a_{\alpha,0}(j) = 1$  for  $\alpha = 0$ , as both sides equal  $x_0$  in this case. Assuming that (4.17) holds for some  $\alpha \geq 0$  with  $a_{\alpha,0}(j) = 1$ , then (4.12) of Lemma 4.3 implies that (4.17) holds for  $\alpha + 1$  and  $\beta = 0$  with

$$a_{\alpha+1,0}(j) = \frac{b-j}{b} a_{\alpha,0}(j) + \frac{j}{b} a_{\alpha,0}(j-1) = 1.$$

Hence, (4.17) holds for all  $\alpha \geq 0$  and  $\beta = 0$  with  $a_{\alpha,0}(j) = 1$ . Similarly, assuming now that (4.17) holds for  $a_{\alpha,\beta-1}(j)$  with nonnegative  $\alpha, \beta - 1$ , (4.17) holds with  $a_{\alpha,\beta}(j)$  given by (4.13) of Lemma 4.3, thus completing the induction.

As our computations involve only moments up to fourth order, we highlight these particular special cases of Lemma 4.4 in the following corollary.

**Corollary 4.1.** For  $\alpha, \beta \geq 0$ , the probability  $g_{\alpha,\beta}$  in (3.6) is given by

$$g_{\alpha,\beta} = f_{\alpha,\beta,n_{n/2}} p_{\alpha,\beta}, \text{ where } p_{\alpha,\beta} = \frac{(n/2)_\alpha (n/2)_\beta}{(n)_{\alpha+\beta}}, \tag{4.18}$$

and  $f_{\alpha,\beta,n_{n/2}}$  is given by (4.15). For  $0 \leq \alpha + \beta \leq 4$ , the sequences in (4.15) specialize to

$$\begin{aligned} a_{0,0}(j) &= 1, \\ a_{0,1}(j) &= (-1)^j \text{ and } a_{1,0}(j) = 1, \\ a_{0,2}(j) &= (-1)^j, \quad a_{1,1}(j) = 1 - j, \text{ and } a_{2,0}(j) = 1, \\ a_{0,3}(j) &= (-1)^j, \quad a_{1,2}(j) = (-1)^j \left(1 - \frac{2j}{3}\right), \\ a_{2,1} &= 1 - \frac{2j}{3}, \text{ and } a_{3,0}(j) = 1, \end{aligned}$$

and

$$a_{0,4}(j) = (-1)^j, \quad a_{1,3}(j) = (-1)^j \left(1 - \frac{j}{2}\right), \quad a_{2,2} = 1 - \frac{j(4-j)}{3},$$

$$a_{3,1} = 1 - \frac{1}{2}j, \quad \text{and} \quad a_{4,0}(j) = 1.$$

*Proof.* By Lemma 4.2, that is, the fact that the switch variables can be applied in any order, conditioning on the values of the switch variables in stage  $n/2$  yields the same result as assuming these values as initial conditions in stage 0, and applying the switch pattern  $\mathbf{n}_{n/2}$ , that is,  $\mathbf{n}$  skipping stage  $n/2$ . Hence, the first claim in (4.18) follows, as the first factor is the probability of the given event conditioned on the values in stage  $n/2$ , while the second factor is the probability of the conditioning event, as

$$P(X_{n/2,1} = \dots = X_{n/2,\alpha} = 0, X_{n/2,\alpha+1} = \dots = X_{n/2,\alpha+\beta} = 1)$$

$$= \prod_{i=0}^{\alpha-1} \binom{n/2-i}{n-i} \prod_{i=0}^{\beta-1} \binom{n/2-i}{n-\alpha-i},$$

which is  $p_{\alpha,\beta}$ .

The specific forms of the sequences  $a_{\alpha,\beta}(j)$  for  $0 \leq \alpha + \beta \leq 4$  follow directly from the initial condition and recursion in Lemma 4.4.

Applying Corollary 4.1, we obtain, for example, the formulae

$$g_{2,1} = \frac{1}{8} (1 + \lambda_{n,1,\mathbf{n}_{n/2}} - \lambda_{n,2,\mathbf{n}_{n/2}} - \lambda_{n,3,\mathbf{n}_{n/2}}) \frac{(n/2)_2 n/2}{(n)_3}$$

and

$$g_{2,2} = \frac{1}{16} (1 - 2\lambda_{n,2,\mathbf{n}_{n/2}} + \lambda_{n,4,\mathbf{n}_{n/2}}) \frac{(n/2)_2 (n/2)_2}{(n)_4}.$$

Lastly, we present the bounds on products of eigenvalues of the chain used to handle the variance term (2.2) when applying Theorem 1.1 to the lightbulb chain.

**Lemma 4.5.** *For all even  $n \geq 6$ ,*

$$|\lambda_{n,2,\mathbf{n}_{n/2}}| \leq e^{-n} \quad \text{and} \quad |\lambda_{n,4,\mathbf{n}_{n/2}}| \leq \frac{1}{2}e^{-n}. \tag{4.19}$$

*Proof.* We show only that the first inequality of (4.19) holds. The remaining claim can be obtained by similar, but more extensive calculations; see the technical report [7] for complete derivations, which slightly generalize the arguments of [8].

Let  $n \geq 2$ . Consider the second-degree polynomial (cf. (1.3))

$$f_2(x) = 1 - \frac{4x}{n} + \frac{4(x)_2}{(n)_2}, \quad 0 \leq x \leq n.$$

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

$$|f_2(x)| \leq \begin{cases} \frac{1}{n-1} & \text{for } x \in \left[ \frac{n-\sqrt{n}}{2}, \frac{n+\sqrt{n}}{2} \right], \\ \left(1 - \frac{2x}{n}\right)^2 & \text{for } x \in \left[ \frac{n-\sqrt{n}}{2}, \frac{n+\sqrt{n}}{2} \right]^c \cap [0, n]. \end{cases} \tag{4.20}$$

For  $x \in \mathbb{R}$ , let  $\lfloor x \rfloor$  and  $\lceil x \rceil$  denote the greatest integer less than or equal to  $x$ , and the smallest integer greater than or equal to  $x$ , respectively. Letting

$$t = \left\{ \left\lfloor \frac{n - \sqrt{n}}{2} \right\rfloor, \dots, \left\lfloor \frac{n + \sqrt{n}}{2} \right\rfloor \right\} \setminus \left\{ \frac{n}{2} \right\},$$

we have

$$|\lambda_{n,2,n_{n/2}}| = \left( \prod_{s=0}^{\lfloor (n-\sqrt{n})/2 \rfloor} |f_2(s)| \right) \left( \prod_{s \in t} |f_2(s)| \right) \left( \prod_{s=\lceil (n+\sqrt{n})/2 \rceil}^n |f_2(s)| \right).$$

If either of the roots  $(n - \sqrt{n})/2$  or  $(n + \sqrt{n})/2$  is an integer then equality holds as both expressions above are 0. Now assuming that neither value is an integer, the product below is over disjoint indices.

Applying the bound (4.20),  $\lfloor n/2 - x \rfloor + \lceil n/2 + x \rceil = n$  and  $1 - 2(n - s)/n = -(1 - 2s/n)$  yields

$$\begin{aligned} |\lambda_{n,2,n_{n/2}}| &\leq \left( \prod_{s=0}^{\lfloor (n-\sqrt{n})/2 \rfloor} \left( 1 - \frac{2s}{n} \right)^2 \right) \left( \prod_{s \in t} \frac{1}{n-1} \right) \left( \prod_{s=\lceil (n+\sqrt{n})/2 \rceil}^n \left( 1 - \frac{2s}{n} \right)^2 \right) \\ &= \left( \prod_{s=0}^{\lfloor (n-\sqrt{n})/2 \rfloor} \left( 1 - \frac{2s}{n} \right) \right)^4 \left( \frac{1}{n-1} \right)^{|t|}, \end{aligned}$$

where  $|t|$  is the cardinality of  $t$ .

Using  $1 - x \leq e^{-x}$  and the fact that  $\lfloor x \rfloor \geq x - 1$  on the first product, we obtain the bound

$$\begin{aligned} |\lambda_{n,2,n_{n/2}}| &\leq \left( \exp \left[ -\frac{2}{n} \left( \left\lfloor \frac{n - \sqrt{n}}{2} \right\rfloor \right) \left( \left\lfloor \frac{n - \sqrt{n}}{2} \right\rfloor + 1 \right) / 2 \right] \right)^4 e^{-|t| \log(n-1)} \\ &\leq \exp \left[ -\left( n - 2\sqrt{n} - 1 + \frac{2}{\sqrt{n}} + |t| \log(n-1) \right) \right]. \end{aligned} \tag{4.21}$$

To control  $|t|$ , note that, as  $\lceil x \rceil \leq x + 1$ , 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

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

and, hence, from (4.21),

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

It can be verified directly that  $\lambda_{n,2,n_{n/2}}$  satisfies this same bound for all even integers  $6 \leq n \leq 34$ , thus completing the proof.

## References

- [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*, eds T. W. Anderson, K. B. Athreya and D. L. Iglehart, Academic Press, Boston, MA, pp. 59–81.
- [2] CHEN, L. H. Y., GOLDSTEIN, L. AND SHAO, Q. M. (2010). *Normal Approximation by Stein's Method*. Springer, Heidelberg.
- [3] GHOSH, S. AND GOLDSTEIN, L. (2011). Concentration of measures via size-biased couplings. *Prob. Theory Relat. Fields* **149**, 271–278.
- [4] GOLDSTEIN, L. (2005). Berry-Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing. *J. Appl. Prob.* **42**, 661–683.
- [5] GOLDSTEIN, L. AND PENROSE, M. D. (2010). Normal approximation for coverage models over binomial point processes. *Ann. Appl. Prob.* **20**, 696–721.
- [6] GOLDSTEIN, L. AND RINOTT, Y. (1996). Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Prob.* **33**, 1–17.
- [7] GOLDSTEIN, L. AND ZHANG, H. (2011). A Berry Esseen theorem for the lightbulb process. Preprint. Available at <http://arxiv.org/abs/1001.0612v3>.
- [8] RADHAKRISHNA, R. C., BHASKARA, R. M. AND ZHANG, H. (2007). One bulb? Two bulbs? How many bulbs light up?—a discrete probability problem involving dermal patches. *Sankyā* **69**, 137–161.
- [9] SHAO, Q. M. AND SU, Z. (2005). The Berry-Esseen bound for character ratios. *Proc. Amer. Math. Soc.* **134**, 2153–2159.
- [10] STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9**, 1135–1151.
- [11] STEIN, C. (1986). *Approximate Computation of Expectations*. Institute of Mathematical Statistics, Hayward, CA.
- [12] ZHOU, H. AND LANGE, K. (2009). Composition Markov chains of multinomial type. *Adv. Appl. Prob.* **41**, 270–291.