On the Number of Switches in Unbiased Coin-tossing

Wenbo V. Li*
Department of Mathematical Sciences
University of Delaware
wli@math.udel.edu

January 25, 2013

Abstract

A biased coin is tossed n times independently and sequentially. A "head" switch is a tail followed by a head and a "tail" switch is a head followed by a tail. Joint Laplace transform for the number of "head" switches and "tail" switches are given. For the total number of switches, the central limit theorem and the large deviation principle are established.

1 Introduction

The following question was posed by Anush (2012) at mathoverflow, asking for bounds for number of coin toss switches: "I toss n biased coins and I want to count the number of times you get a H followed by a T or a T followed by a H. I call these switches. So for example if I get $HHTTHTHHHT$ then I have 5 switches in total. If the coin gives H with probability p and T with probability $1-p$ then how can you find an approximation to the probability of getting at least k switches for large n? I would also be interested in a Chernoff style tail bound."

Let S_n denote the number of switches. As it is pointed out in Anush (2012) that adjacent switch occurrences are not independent however non-adjacent ones are. And the probability of having a switch at position $i+1$ given that there is a switch at position i is exactly $1/2$, irrespective of p. The mean number of switches is $\mathbb{E}\, S_n = (n-1)2pq$ and the variance is $\mathrm{Var}(S_n) = 2pq(2n-3-2pq(3n-5))$ where $q = (1-p)$. The exact probability distribution for S_n was given by Joriki (2012). Namely

$$\mathbb{P}(S_n = 2l-1) = 2\sum_{j=l}^{n-l} \binom{j-1}{l-1}\binom{n-j-1}{l-1} p^j (1-p)^{n-j},$$

*supported by NSF grant DMS-1106938

for $1 \le l \le \lfloor n/2 \rfloor$, and

$$\mathbb{P}(S_n = 2l) = \sum_{j=l}^{n-l-1} \binom{j-1}{l-1}\binom{n-j-1}{l}\left(p^j(1-p)^{n-j} + p^{n-j}(1-p)^j\right) ,$$

for $0 \le l \le \lfloor (n-1)/2 \rfloor$. Ori Gurel-Gurevich pointed out in mathoverflow that a Chernoff style bound can be given easily by considering the "even" and "odd" switches separately. The number of even (odd) switches is binomially distributed and one can use union bound and lose at most a factor of 2 in the bound. However, this is one-sided and not sharp, see remarks after Theorem 2. Ofir also mentioned that one may try recursively compute $f(n,k,p)$, the probability that during n tosses there are exactly k switches and that the last result is H, by using $f(n+1,k,p) = p(f(n,k,p) + f(n,k-1,1-p))$ with $f(n,0,p) = p^n$. Our approach is different and treat the last flip together with the rest.

The main propose of this note is finding explicitly the generating function or Laplace transform $\mathbb{E}\, e^{\lambda S_n}$, $\mathbb{E}\, e^{\lambda S_n(H)}$ and $\mathbb{E}\, e^{\lambda S_n(T)}$, where $S_n = S_n(H) + S_n(T)$, $S_n(H)$ is the number of "head" switches from T to H and $S_n(T)$ is the number of "tail" switches from H to T. The crucial observation is the representations

$$S_n(H) = \sum_{i=2}^{n}(1-\varepsilon_{i-1})\varepsilon_i \quad \text{and} \quad S_n(T) = \sum_{i=2}^{n}\varepsilon_{i-1}(1-\varepsilon_i) \tag{1.1}$$

where ε_i are i.i.d and $\mathbb{P}(\varepsilon_i = 1) = p, \mathbb{P}(\varepsilon_i = 0) = 1 - p = q$. Our detailed analysis of using two related generating functions is motivated from an analytic argument by the author a few years ago to the following neat fact mentioned by Persi Diaconis: For independent random variables X_i with $\mathbb{P}(X_i = 1) = 1/i = 1 - \mathbb{P}(X_i = 0)$, $i \ge 1$,

$$\sum_{j=1}^{\infty} X_j X_{j+1} =^d \text{Poisson}(1). \tag{1.2}$$

We include an outline of the proof for the above fact in Section 2 after the proof of the following main result of this note.

Theorem 1 *For any $\lambda, \eta \in \mathbb{R}$, the joint Laplace transform*

$$\begin{aligned}\mathbb{E}\, e^{\lambda S_n(H)+\eta S_n(T)} &= \frac{1}{2}\left(1 + \frac{1-4pq+2pq(e^{\lambda}+e^{\eta})}{\sqrt{1-4pq+4pqe^{\lambda+\eta}}}\right) r_1^{n-1} \\ &+ \frac{1}{2}\left(1 - \frac{1-4pq+2pq(e^{\lambda}+e^{\eta})}{\sqrt{1-4pq+4pqe^{\lambda+\eta}}}\right) r_2^{n-1}\end{aligned} \tag{1.3}$$

where

$$r_{1,2} = \frac{1}{2}(1 \pm \sqrt{(1-4pq)+4pqe^{\lambda+\eta}}). \tag{1.4}$$

We need several remarks. First note that $r_2^2 < r_1^2$ (r_1 is defined with the positive sign and r_2 negative) and so the dominating term in (1.3) is r_1^n. Second, our technique used in the proof also allow us to find the joint Laplace transform for $S_n(HH), S_n(HT) = S_n(T), S_n(TH) = S_n(H), S_n(TT)$ where $S_n(F_1F_2)$ denote the number of consecutive pattens F_1F_2 in the first n flips, with $F_i \in \{H,T\}$. Indeed, one could obtain joint Laplace transform for 2^k correlated variables $\{S_n(F_1 \cdots F_k) : F_i \in \{H,T\}, 1 \le i \le k\}$ for any fixed $k \ge 2$. But we will not consider these generalizations in this note. Third, it follows from (1.3) that $S_n(H) =^d S_n(T)$ in distribution. We can also find easily from Theorem 1 or its proof in Section 2 that $\mathbb{E}\, S_n(H) = \mathbb{E}\, S_n(T) = pq(n-1) \sim pqn$,

$$\mathrm{Var}(S_n(H)) = \mathrm{Var}(S_n(T)) = pq(1-3pq)n - pq(1-5pq) \sim pq(1-3pq)n, \tag{1.5}$$

and

$$\mathrm{Cov}(S_n(H), S_n(T)) = pq(1-3pq)n - pq(2-5pq) \sim pq(1-3pq)n. \tag{1.6}$$

Next, to obtain useful information about numbers of various types of switches, we have the following CLTs and deviation estimates.

Theorem 2 *For $0 < p < 1$, as $n \to \infty$, we have the joint CLT*

$$\left(\frac{S_n(H) - pqn}{\sqrt{pq(1-3pq)n}}, \frac{S_n(T) - pqn}{\sqrt{pq(1-3pq)n}} \right) \Rightarrow N\left((0,0); \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right) \tag{1.7}$$

and as a consequence for $S_n = S_n(H) + S_n(T)$,

$$\frac{S_n - 2pqn}{\sqrt{4pq(1-3pq)n}} \Rightarrow N(0,1). \tag{1.8}$$

In addition, the large deviation principle holds for S_n/n with convex rate function

$$\begin{aligned} \Lambda^*(\theta) &= \frac{\theta}{2} \log \left(\frac{2\theta(1-4pq)}{\sqrt{\theta^2 + 4(1-\theta)(1-4pq)} - 4pq\theta - (1-4pq)(2-\theta)} \right) \\ &\quad + \log \left(\frac{2 - \theta - \sqrt{\theta^2 + 4(1-\theta)(1-4pq)}}{4pq} \right), \quad 0 < \theta < 1, \end{aligned}$$

and $\Lambda^*(\theta) = \infty$ otherwise.

Note that in particular, we have for $\rho > 2pq$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \ge \rho n) = -\Lambda^*(\rho),$$

and for $0 < \rho < 2pq$

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \le \rho n) = -\Lambda^*(\rho).$$

As pointed out by Ori Gurel-Gurevich mentioned earlier, a Chernoff style bound can be given easily by considering the "even" and "odd" switches separately. However, this is one-sided and not sharp in comparison with estimates above.

The remaining of this note is organized as follows. The joint Laplace transforms for numbers of head and tail switches are given in section 2 with the help of two related generating functions. Similar idea is used to establish (1.2) at the end of section 2. The proof of Theorem 2 for CLT and large deviation principle are given in section 3.

2 Joint Transforms for Head and Tail Switches

Using notations and representations (1.1) in the introduction, We have by conditioning on ε_n for $n \geq 2$, the joint Laplace transform

$$\begin{aligned} \mathbb{E}\, e^{\lambda S_n(H)+\eta S_n(T)} &= p\mathbb{E}\, e^{\lambda S_{n-1}(H)+\eta S_{n-1}(T)+\lambda(1-\varepsilon_{n-1})} + q\mathbb{E}\, e^{\lambda S_{n-1}(H)+\eta S_{n-1}(T)+\eta\varepsilon_{n-1}} \\ &= pa_{n-1}(\lambda,\eta) + qb_{n-1}(\lambda,\eta). \end{aligned} \tag{2.9}$$

Here $a_0(\lambda,\eta) = 1$, $a_1(\lambda,\eta) = p + qe^{\lambda}$,

$$\begin{aligned} a_n(\lambda,\eta) &:= \mathbb{E}\, e^{\lambda S_n(H)+\eta S_n(T)+\lambda(1-\varepsilon_n)} \\ &= p\mathbb{E}\, e^{\lambda S_{n-1}+\lambda(1-\varepsilon_{n-1})+\eta S_{n-1}(T)} + qe^{\lambda}\mathbb{E}\, e^{\lambda S_{n-1}(H)+\eta S_{n-1}(T)+\eta\varepsilon_{n-1}} \\ &= pa_{n-1}(\lambda,\eta) + qe^{\lambda} b_{n-1}(\lambda,\eta), \quad n \geq 1 \end{aligned} \tag{2.10}$$

and similarly, $b_0(\lambda) = 1$, $b_1(\lambda,\eta) = pe^{\eta} + q$,

$$\begin{aligned} b_n(\lambda,\eta) &:= \mathbb{E}\, e^{\lambda S_n(H)+\eta S_n(T)+\eta\varepsilon_n} \\ &= pe^{\eta} a_{n-1}(\lambda,\eta) + qb_{n-1}(\lambda,\eta), \quad n \geq 1. \end{aligned} \tag{2.11}$$

Consider generating functions

$$A_{\lambda,\eta}(x) := \sum_{n\geq 0} a_n(\lambda,\eta)x^n, \quad B_{\lambda,\eta}(x) := \sum_{n\geq 0} b_n(\lambda,\eta)x^n.$$

Then from (2.10) and (2.11),

$$\begin{aligned} A_{\lambda,\eta}(x) &= 1 + px\sum_{n\geq 1} a_{n-1}x^{n-1} + qe^{\lambda}x\sum_{n\geq 1} b_{n-1}x^{n-1} \\ &= 1 + pxA_{\lambda,\eta}(x) + qe^{\lambda}xB_{\lambda,\eta}(x), \\ B_{\lambda,\eta}(x) &= 1 + pe^{\lambda}x\sum_{n\geq 1} a_{n-1}x^{n-1} + qx\sum_{n\geq 1} b_{n-1}x^{n-1} \\ &= 1 + pe^{\lambda}xA_{\lambda,\eta}(x) + qxB_{\lambda,\eta}(x). \end{aligned}$$

Solving the above system of equations above and we see

$$A_{\lambda,\eta}(x) = \sum_{n\geq 0} a_n(\lambda,\eta)x^n = \frac{1 + q(e^{\lambda}-1)x}{1 - x - pq(e^{\lambda+\eta}-1)x^2}$$

and (also by symmetry)

$$B_{\lambda,\eta}(x) = \sum_{n\geq 0} b_n(\lambda,\eta)x^n = \frac{1+p(e^{\eta}-1)x}{1-x-pq(e^{\lambda+\eta}-1)x^2}.$$

Thus we find by partial fraction,

$$A_{\lambda,\eta}(x) = \frac{1+q(e^{\lambda}-1)x}{1-x-pq(e^{2\lambda}-1)x^2} = \frac{C_1}{1-r_1x} + \frac{C_2}{1-r_2x} \tag{2.12}$$

where r_1 and r_2 are defined in (1.4), and

$$C_{1,2} = \frac{1}{2}\left(1 \pm \frac{1-2q+2qe^{\lambda}}{\sqrt{(1-4pq)+4pqe^{\lambda+\eta}}}\right). \tag{2.13}$$

Thus we have from (2.12)

$$a_n(\lambda,\eta) = C_1r_1^n + C_2r_2^n$$

and by symmetry (exchange p and q, λ and η)

$$b_n(\lambda,\eta) = \frac{1}{2}\left(1 + \frac{1-2p+2pe^{\eta}}{\sqrt{(1-4pq)+4pqe^{\lambda+\eta}}}\right)r_1^n + \frac{1}{2}\left(1 - \frac{1-2p+2pe^{\eta}}{\sqrt{(1-4pq)+4pqe^{\lambda+\eta}}}\right)r_2^n.$$

Substituting into (2.9), we obtain (1.3) and finish the proof of Theorem 1. □

As mentioned in the introduction, our technique used above allow us to find the joint Laplace transform for $S_n(HH), S_n(HT) = S_n(T), S_n(TH) = S_n(H), S_n(TT)$ where $S_n(F_1F_2)$ denote the number of consecutive patterns F_1F_2 in the first n flips, with $F_i \in \{H,T\}$. However, we will not do it in this note and instead we prefer to show (1.2) here. To see a formal connection with (1.2), we note that the number of non-switches can be represented as

$$S_n(HH) = \sum_{i=2}^{n} \varepsilon_{i-1}\varepsilon_i \quad \text{and} \quad S_n(TT) = \sum_{i=2}^{n}(1-\varepsilon_{i-1})(1-\varepsilon_i). \tag{2.14}$$

Define for independent X_i, $i \geq 1$

$$T_n = \sum_{j=1}^{n} X_jX_{j+1}, \qquad \text{with} \quad \mathbb{P}(X_i = 1) = 1/i = 1 - \mathbb{P}(X_i = 0).$$

Then by conditioning, the characteristic function

$$\begin{aligned}\phi_n(t) = \mathbb{E}\, e^{\mathrm{i}tT_n} &= \mathbb{P}(X_{n+1}=0)\cdot\phi_{n-1}(t) + \mathbb{P}(X_{n+1}=1)\cdot\mathbb{E}\,(e^{\mathrm{i}tT_n}|X_{n+1}=1) \\ &= \mathbb{P}(X_{n+1}=0)\cdot\phi_{n-1}(t) + \mathbb{P}(X_{n+1}=1)\cdot g_n(t)\end{aligned}$$

where

$$\begin{aligned} g_n(t) &= \mathbb{E}\,(e^{\mathtt{i}tT_n}|X_{n+1}=1) \\ &= \mathbb{P}(X_n=0)\cdot\mathbb{E}\,(e^{\mathtt{i}tT_n}|X_n=0,X_{n+1}=1)+\mathbb{P}(X_n=1)\cdot\mathbb{E}\,(e^{\mathtt{i}tT_n}|X_n=X_{n+1}=1) \\ &= \mathbb{P}(X_n=0)\cdot\phi_{n-2}(t)+\mathbb{P}(X_n=1)\cdot e^{\mathtt{i}t}\cdot g_{n-1}(t). \end{aligned}$$

Thus we have two recursive function relations:

$$\begin{aligned} g_n(t) &= (n+1)\phi_n(t)-n\phi_{n-1}(t) \\ g_n(t) &= \frac{e^{\mathtt{i}t}}{n}g_{n-1}(t)+\frac{n-1}{n}\phi_{n-2}(t) \end{aligned}$$

Substituting the first one into the second one, we have

$$\phi_n(t)=\phi_{n-1}(t)+\frac{a}{n+1}\phi_{n-1}-\frac{(n-1)a}{n(n+1)}\phi_{n-2}$$

with $a=e^{\mathtt{i}t}-1$ and the initial condition

$$\begin{aligned} \phi_1 &= \frac{1}{2}+\frac{1}{2}e^{\mathtt{i}t}=1+\frac{1}{2}a \\ \phi_2 &= \frac{1}{2}+\frac{1}{3}e^{\mathtt{i}t}+\frac{1}{6}e^{2\mathtt{i}t}=1+\frac{2}{3}a+\frac{1}{6}a^2. \end{aligned}$$

It can be easily proved by mathematical induction that

$$\phi_n=1+\sum_{k=1}^{n}\left(\frac{1}{k!}-\frac{1}{(n+1)(k-1)!}\right)a^k.$$

Taking $n\to\infty$, we obtain

$$\phi_n(t)=\mathbb{E}\,e^{\mathtt{i}tT_n}\to e^a=e^{e^{\mathtt{i}t}-1}$$

which proves (1.2). □

3 Proof of Theorem 2

From the key representation (1.1), we see easily $\mathbb{E}\,S_n(H)=\mathbb{E}\,S_n(T)=pq(n-1)\sim pqn$ and

$$\begin{aligned} \mathrm{Var}(S_n(H)) &= \mathbb{E}\,(S_n(H)-\mathbb{E}\,S_n(H))^2 \\ &= \sum_{i=2}^{n}\mathbb{E}\,(\varepsilon_i-\varepsilon_{i-1}\varepsilon_i-pq)^2+2\sum_{i=2}^{n-1}\mathbb{E}\,(\varepsilon_i-\varepsilon_{i-1}\varepsilon_i-pq)(\varepsilon_{i+1}-\varepsilon_i\varepsilon_{i+1}-pq) \\ &= pq(1-pq)(n-1)-2p^2q^2(n-2) \\ &= pq(1-3pq)n-pq(1-5qp)\sim pq(1-3pq)n. \end{aligned}$$

Similarly, we have (1.5) and (1.6). Thus from Theorem 1, for any $\lambda, \eta > 0$, as $n \to \infty$,

$$\begin{aligned}\Psi_n(\lambda,\eta) &= \mathbb{E}\exp\left(\lambda\frac{S_n(H)-\mathbb{E}\,S_n(H)}{\sqrt{\mathrm{Var}(S_n(H))}}+\eta\frac{S_n(T)-\mathbb{E}\,S_n(T)}{\sqrt{\mathrm{Var}(S_n(T))}}\right)\\ &= \mathbb{E}\exp\left(\lambda\frac{S_n(H)-pqn}{\sqrt{pq(1-3pq)n}}+\eta\frac{S_n(T)-pqn}{\sqrt{pq(1-3pq)n}}+o(1)\right)\\ &= \exp\left(n\log\frac{1}{2}(1+\sqrt{\Delta})-\frac{(\lambda+\eta)pq\sqrt{n}}{\sqrt{pq(1-3pq)}}+o(1)\right) \qquad (3.15)\end{aligned}$$

where from (1.4),

$$\Delta = (1-4pq)+4pq\exp\left((\lambda+\eta)/\sqrt{pq(1-3pq)n}\right).$$

Note that as $n \to \infty$,

$$\begin{aligned}\Delta - 1 &= 4pq\left(\exp\left((\lambda+\eta)/\sqrt{pq(1-3pq)n}\right)-1\right)\\ &= \frac{4pq(\lambda+\eta)}{\sqrt{pq(1-3pq)n}}+\frac{2(\lambda+\eta)^2}{(1-3pq)n}+o(\frac{1}{n}).\end{aligned}$$

Hence as $n \to \infty$

$$\begin{aligned}\frac{1}{2}(1+\sqrt{\Delta}) &= 1+\frac{1}{4}(\Delta-1)-\frac{1}{16}(\Delta-1)^2+o(\frac{1}{n})\\ &= 1+\frac{pq(\lambda+\eta)}{\sqrt{pq(1-3pq)n}}+\frac{(1-2pq)(\lambda+\eta)^2}{2(1-3pq)n}+o(\frac{1}{n})\end{aligned}$$

and

$$\begin{aligned}\log\frac{1}{2}(1+\sqrt{\Delta}) &= \frac{pq(\lambda+\eta)}{\sqrt{pq(1-3pq)n}}+\frac{(1-2pq)(\lambda+\eta)^2}{2(1-3pq)n}+\frac{1}{2}\cdot\frac{pq(\lambda+\eta)^2}{(1-3pq)n}+o(\frac{1}{n})\\ &= \frac{pq(\lambda+\eta)}{\sqrt{pq(1-3pq)n}}+\frac{(\lambda+\eta)^2}{2n}+o(\frac{1}{n}). \qquad (3.16)\end{aligned}$$

Combining (3.15) and (3.16), we obtain

$$\Psi_n(\lambda,\eta) = \mathbb{E}\exp\left(\lambda\frac{S_n(H)-\mathbb{E}\,S_n(H)}{\sqrt{\mathrm{Var}(S_n(H))}}+\eta\frac{S_n(T)-\mathbb{E}\,S_n(T)}{\sqrt{\mathrm{Var}(S_n(T))}}\right) \sim e^{\lambda^2/2}, \quad \text{as} \quad n \to \infty,$$

which implies the joint CLT in Theorem 2.

For the large deviation principle, we need to compute the logarithmic generating function for $S_n = S_n(H) + S_n(T)$. From Theorem 1 with $\eta = \lambda$, the logarithmic generating function for S_n is

$$\begin{aligned}\Lambda(\lambda) &= \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\, \exp(\lambda S_n) = \log r_1 \\ &= -\log 2 + \log\left(1 + \sqrt{1 - 4pq + 4pqe^{2\lambda}}\right)\end{aligned}$$

and the associated Fenchel-Legendre transform of $\Lambda(\lambda)$ is

$$\Lambda^*(\theta) = \sup_{\lambda\in\mathbb{R}} \left(\theta\lambda - \Lambda(\lambda)\right).$$

Since $0 < \Lambda'(\lambda) < 1$, we see $\Lambda^*(\theta) = \infty$ for $\theta \notin (0, 1)$. For $0 < \theta < 1$, we set λ_θ such that

$$\theta = \Lambda'(\lambda_\theta) = \frac{4pqe^{2\lambda_\theta}}{1 - 4pq + 4pqe^{2\lambda_\theta} + \sqrt{1 - 4pq + 4pqe^{2\lambda_\theta}}}.$$

Then solving the resulting quadratic equation, we find

$$\begin{aligned}\sqrt{1 - 4pq + 4pqe^{2\lambda_\theta}} &= \frac{\theta + \sqrt{\theta^2 + 4(1-\theta)(1-4pq)}}{2(1-\theta)} \\ &= 2(1-4pq)\left(\sqrt{\theta^2 + 4(1-\theta)(1-4pq)} - \theta\right)^{-1}.\end{aligned}$$

Thus for $0 < \theta < 1$, by substituting λ_θ from the equation above,

$$\begin{aligned}\Lambda^*(\theta) &= \theta\lambda_\theta - \Lambda(\lambda_\theta) \qquad (3.17)\\ &= \frac{\theta}{2}\log\left(\frac{2\theta(1-4pq)}{\sqrt{\theta^2 + 4(1-\theta)(1-4pq)} - 4pq\theta - (1-4pq)(2-\theta)}\right) \\ &\quad + \log\left(\frac{2 - \theta - \sqrt{\theta^2 + 4(1-\theta)(1-4pq)}}{4pq}\right). \qquad (3.18)\end{aligned}$$

Note that $\Lambda^*(2pq) = 0$ as it should be. By the Gartner-Ellis theorem, see [2], the large deviation principle holds for S_n/n with rate function $\Lambda^*(\theta)$ since $\Lambda(\lambda)$ exist for all $\lambda \in \mathbb{R}$ and is essentially smooth. In particular, we have for $\rho > 2pq$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq \rho n) = -\inf_{\theta\geq\rho} \Lambda^*(\theta) = -\Lambda^*(\rho),$$

and for $0 < \rho < 2pq$

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \leq \rho n) = -\inf_{\theta\leq\rho} \Lambda^*(\theta) = -\Lambda^*(\rho).$$

Finally, it is also easy to check from (3.18) that

$$\lim_{p\to 1/2} \Lambda^*(\theta) = \theta \log(2\theta) + (1-\theta)\log(2(1-\theta))$$

which is the rate function for binomial random variable S_n when $p = 1/2$.

References

[1] Anush (2012), Bounds for number of coin tosss witches, http://mathoverflow.net/questions/116452

[2] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Springer-Verlag, Berlin, 2010.

[3] P. Diaconis (2010), Personal comunication.

[4] Joriki (2012), http://math.stackexchange.com/questions/258221