ROCKY MOUNTAIN
JOURNAL OF MATHEMATICS
Volume 48, Number 6, 2018

WHEN FOURTH MOMENTS ARE ENOUGH

CHRIS JENNINGS-SHAFFER, DANE R. SKINNER
AND EDWARD C. WAYMIRE

ABSTRACT. This note concerns a somewhat innocent question motivated by an observation concerning the use of Chebyshev bounds on sample estimates of p in the binomial distribution with parameters n, p, namely, what moment order produces the best Chebyshev estimate of p? If $S_n(p)$ has a binomial distribution with parameters n, p, then it is readily observed that $\text{argmax}_{0\le p\le 1}\mathbb{E}S_n^2(p) = \text{argmax}_{0\le p\le 1}np(1-p) = 1/2$, and $\mathbb{E}S_n^2(1/2) = n/4$. Bhattacharya **[2]** observed that, while the second moment Chebyshev sample size for a 95 percent confidence estimate within ± 5 percentage points is $n = 2000$, the fourth moment yields the substantially reduced polling requirement of $n = 775$. Why stop at the fourth moment? Is the argmax achieved at $p = 1/2$ for higher order moments, and, if so, does it help in computing $\mathbb{E}S_n^{2m}(1/2)$? As captured by the title of this note, answers to these questions lead to a simple rule of thumb for the best choice of moments in terms of an effective sample size for Chebyshev concentration inequalities.

1. Introduction. This note concerns a somewhat innocent question motivated by an observation concerning the use of Chebyshev bounds on sample estimates of p in the binomial distribution with parameters n, p, namely, what moment order produces the best Chebyshev estimate of p? Chebyshev is arguably the most basic concentration inequality to occur in risk probability estimates, and the use of second moments is a textbook example in elementary probability and statistics. Consider iid Bernoulli $0-1$ random variables $X_1, X_2, \ldots, X_n$ with parameter $p \in [0,1]$, and let $S_n(p) = \sum_{j=1}^n (X_j - p)$. Then, it is readily observed that $\text{argmax}_{0\le p\le 1}\mathbb{E}S_n^2(p) = \text{argmax}_{0\le p\le 1}np(1-p) = 1/2$. It is also a well-

2010 AMS *Mathematics subject classification.* Primary 60A10, 62D05.

Keywords and phrases. Binomial distribution, estimation, concentration inequalities, machine learning.

This work was partially supported by DMS-1408947, and by the National Science Foundation grant DMS-1408947.

Received by the editors on November 21, 2017, and in revised form on December 23, 2017.

DOI:10.1216/RMJ-2018-48-6-1917

known probability exercise to check that fourth moment Chebyshev bounds improve the rate of convergence that can more generally be used for a proof of the strong law of large numbers, e.g., see [**2,** page 100]. Somewhat relatedly, Bhattacharya [**1**] recently noticed, after a mildly tedious calculation for checking $\operatorname{argmax}_{0\le p\le 1}\mathbb{E}S_n^4(p) = 1/2$, that the second moment Chebyshev bound is rather significantly improved by consideration of fourth moments as well. In particular, while the second moment Chebyshev sample size for a 95 percent confidence estimate within ± 5 percentage points is $n = 2000$, the fourth moment yields the substantially reduced polling requirement of $n = 775$. While the Chebyshev inequality is one among several inequalities used to obtain sample estimates, it is no doubt the simplest; see [**2**] for comparison of fourth order Chebyshev to other concentration inequality bounds and [**4**] for numerical comparisons to higher order Chebyshev bounds.

So why stop at fourth moments? Is $\operatorname{argmax}_{0\le p\le 1}\mathbb{E}S_n^{2m}(p) = 1/2$ for all m, n, and, if so, does it improve the estimate? Somewhat surprisingly we were unable to find a resolution of such basic questions in the published literature. In any case, with the argmax question resolved in part (a) of the next theorem, part (b) provides a direct computation of $\mathbb{E}S_n^{2m}(1/2)$. Part (c) then provides a more readily computable version.

Theorem 1.1.

(a) *For all $m \ge 1$ and n sufficiently large,* $\operatorname{argmax}_{0\le p\le 1}\mathbb{E}S_n^{2m}(p) = 1/2$.

(b) *For all positive m and n,*

$$\mathbb{E}S_n^{2m}\left(\frac{1}{2}\right) = 4^{-m}\sum_{\substack{\mu\in\pi(m)\\ |\mu|\le m\wedge n}}\binom{2m}{2\mu_1,\ldots,2\mu_{|\mu|}}\binom{n}{|\mu|}.$$

(c) *For all positive m and n,*

$$\mathbb{E}S_n^{2m}\left(\frac{1}{2}\right) = 2^{-2m-n}\sum_{k=0}^{n}\binom{n}{k}(2k-n)^{2m}.$$

Here, $\pi(m)$ is the set of ordered integer partitions of m, also referred to as integer compositions, and $|\mu|$ denotes the number of parts of $\mu\in\pi(m)$. We refer to $|\mu|$ as the size *of the partition μ.*

The equivalent calculus challenge is to show, for fixed m, that, for all sufficiently large n,

$$\text{argmax}_{0\le p\le 1}\frac{d^{2m}}{dt^{2m}}(pe^{qt}+qe^{-pt})^n\Big|_{t=0}=\frac{1}{2}. \tag{1.1}$$

The example below illustrates the challenge in locating absolute maxima for such polynomials (in p), especially for proofs by mathematical induction. The proof given here is based on explicit combinatorial computation of $\mathbb{E}S_n^{2m}(p)$ in terms of ordered partitions of $2m$, after introducing a few preliminary lemmas. The lemmas are relatively easy to verify using statistical independence and identical distributions of the terms $X_i - p$ and $X_j - p$, $i \neq j$, and make good exercises in calculus, probability, and number theory. However, we first observe that part (a) of the theorem does not hold for $m > n$.

Counterexample to Theorem 1.1 (a) **for** (**small**) $n < m$**.** Observe, for $n = 1$ and $m = 2$, the function

$$\mathbb{E}S_1^4(p) = p - 4p^2 + 6p^3 - 3p^4, \quad 0 \le p \le 1,$$

has a *minimum* at $p = 1/2$, with two local maxima at $1/2 \pm \sqrt{2}/4$. In particular,

$$\text{argmax}_{0\le p\le 1}\mathbb{E}S_1^4(p) = \frac{1}{2} \pm \frac{\sqrt{2}}{4}.$$

Specifically, the polynomial is generally *not* unimodal. Thus, the restriction to sufficiently large n is necessary for Theorem 1.1 (a). There is also the question of how large is sufficiently large. We do not address this here; however, computations suggest a bound along the lines of $m \le c \cdot n^{\varepsilon}$, with ε a little less than 1/2. We let m_n denote the largest value of m, dependent on n, such that Theorem 1.1 (a) holds for all $m \le m_n$. We leave this as an open problem for determining an exact formula for m_n, as well as determining a formula for $\text{argmax}_{0\le p\le 1}\mathbb{E}_n^{2m}(p)$, $m > m_n$.

2. Proofs and remarks. Let $\pi(2m)$ denote the set of ordered partitions of $2m$. We will use $|\mu| = k$ to denote the number of parts of μ. Finally, for $\mu \in \pi(2m)$, let

$$f_i(\mu, p) = pq^{\mu_i} + q(-p)^{\mu_i},$$
$$0 \le p \le 1, \quad q = (1-p), \quad 1 \le i \le |\mu|.$$

Lemma 2.1. *Let* $0 \le p \le 1$ *and* $q = 1 - p$. *The following hold*:

(a) $S_n(p) \overset{\text{dist}}{=} -S_n(q)$;

(b) $\mathbb{E}S_n^{2m}(p) = \mathbb{E}S_n^{2m}(q)$;

(c) $\mathbb{E}S_n^{2m}(p) = \sum_{\mu \in \pi(2m)} \binom{n}{|\mu|} \binom{2m}{\mu_1, \dots, \mu_{|\mu|}} \prod_{i=1}^{|\mu|} f_i(\mu, p)$;

(d) $\frac{d}{dp}\mathbb{E}S_n^{2m}(p) = \sum_{\mu \in \pi(2m)} \binom{n}{|\mu|} \binom{2m}{\mu_1, \dots, \mu_{|\mu|}} \sum_{i=1}^{|\mu|} f_i'(\mu, p) \prod_{j \ne i}^{|\mu|} f_j(\mu, p)$.

Lemma 2.2. *Let* $\mu \in \pi(2m)$ *and* $1 \le i \le |\mu|$. *Then*:

$$\frac{d}{dp} f_i(\mu, p) = q^{\mu_i}\left(1 - \frac{p}{q}\mu_i\right) + (-1)^{\mu_i + 1} p^{\mu_i}\left(1 - \frac{q}{p}\mu_i\right).$$

It now follows easily that

$$f_i\left(\mu, \frac{1}{2}\right) = \begin{cases} 2^{-\mu_i} & \text{for even } \mu_i, \\ 0 & \text{for odd } \mu_i; \end{cases} \tag{2.1}$$

$$f_i'\left(\mu, \frac{1}{2}\right) = \begin{cases} 0 & \text{for even } \mu_i, \\ -2(\mu_i - 1)2^{-\mu_i} & \text{for odd } \mu_i. \end{cases} \tag{2.2}$$

The keys to the following proof of Theorem 1.1 reside in:

(1) the parity conflicts between (2.1) and (2.2), and

(2) the expansion (d) in Lemma 2.1, viewed as a polynomial in n.

Proof of Theorem 1.1. That $p = 1/2$ is a critical point follows from Lemma 2.1 (d), together with (2.1) and (2.2), by examining the terms $f_i'(\mu, 1/2) \prod_{j \ne i}^{|\mu|} f_j(\mu, 1/2)$. In particular, for partitions of $2m$, if μ_i is odd, then there must be a $j \ne i$ such that μ_j is odd as well. In order to see that $p = 1/2$ is an absolute maximum, the trick is to observe that, for $0 \le p < 1/2 < q$, the leading coefficient of $(d/dp)\,\mathbb{E}S_n^{2m}(p)$, viewed as a polynomial in n, is obtained at the m-part composition, $\mu = (2, 2, \ldots, 2)$ of $2m$, namely, it is obtained from

$$\binom{n}{m}\binom{2m}{2, 2, \ldots, 2} m(q^2 - p^2)(pq)^{m-1}$$

and, therefore, is positive for all $p < 1/2$. Thus, for sufficiently large n,

$$\frac{d}{dp}\mathbb{E}S_n^{2m}(p) > 0 \quad \text{for } 0 \le p < 1/2.$$

In view of the symmetry expressed in Lemma 2.1 (b), it follows that $p = 1/2$ is the unique global maximum.

For Theorem 1.1 (b), simply compute from independence, writing $\widetilde{X}_i = X_i - 1/2$, $i = 1, 2, \ldots, n$. In particular, $\widetilde{X}_i = \pm 1/2$ with equal probabilities. Thus, for $m \ge 1$,

$$\begin{aligned}
\mathbb{E}S_n^{2m}\left(\frac{1}{2}\right) &= \sum_{1 \le j_1, \ldots, j_{2m} \le n} \mathbb{E}\prod_{i=1}^{2m} \widetilde{X}_{j_i} \\
&= \sum_{2m_1 + \cdots + 2m_n = 2m} \prod_{i=1}^{n} \mathbb{E}\widetilde{X}_i^{2m_i} \\
&= \sum_{k=1}^{m \wedge n} \sum_{\substack{2m_1 + \cdots + 2m_n = 2m \\ \#\{j : m_j \ge 1\} = k}} \prod_{i=1}^{n} 4^{-m_i} \\
&= \sum_{k=1}^{m \wedge n} \binom{n}{k} \sum_{\mu = (\mu_1, \ldots, \mu_k) \in \pi(m)} \binom{2m}{2\mu_1, \ldots, 2\mu_k} 4^{-m}.
\end{aligned}$$

Here, we adopt the convention that a sum over an empty set is zero so that, if there are no partitions μ of m with $|\mu| = k$, then the indicated sum is zero for this choice of k. Thus, nonzero contributions to the sum are provided by ordered partitions μ of size $|\mu| \le m \wedge n$.

In order to simplify the computation in terms of ordered partitions (b), we may proceed as follows to obtain the formula in (c). We compute $\mathbb{E}S_n^{2m}(1/2)$ as the $2m$th moment of $S_n(1/2)$, as given in (1.1). By the binomial theorem, we have that

$$\begin{aligned}\mathbb{E}S_n^{2m}\left(\frac{1}{2}\right) &= \frac{d^{2m}}{dt^{2m}}\left[\left(\frac{e^{t/2}}{2}+\frac{e^{-t/2}}{2}\right)^n\right]_{t=0} \\ &= \frac{d^{2m}}{dt^{2m}}\left[2^{-n}\sum_{k=0}^{n}\binom{n}{k}e^{t(2k-n)/2}\right]_{t=0} \\ &= 2^{-n-2m}\sum_{k=0}^{n}\binom{n}{k}(2k-n)^{2m}. \qquad \square\end{aligned}$$

A linear recurrence in m is also possible in aiding the pre-asymptotic (in n) iterative computations of $\mathbb{E}S_n^{2m}(1/2)$, namely,

Proposition 2.3.

$$\mathbb{E}S_n^{2m+2\ell+2}\left(\frac{1}{2}\right) = \sum_{j=0}^{\ell} c_j 2^{2j-2\ell-2}\mathbb{E}S_n^{2m+2j}\left(\frac{1}{2}\right), \tag{2.3}$$

where $\ell = \lfloor (n-1)/2 \rfloor$, $a_k = (2k-n)^2$ *and* $(c_0, c_1, \ldots, c_\ell)$ *is the unique solution to*

$$\begin{pmatrix} a_0^0 & a_0^1 & \cdots & a_0^\ell \\ a_1^0 & a_1^1 & \cdots & a_1^\ell \\ \vdots & & & \\ a_\ell^0 & a_\ell^1 & \cdots & a_\ell^\ell \end{pmatrix}\begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_\ell\end{pmatrix} = \begin{pmatrix} a_0^{\ell+1} \\ a_1^{\ell+1} \\ \vdots \\ a_\ell^{\ell+1}\end{pmatrix}.$$

Proof. In order to see this, write

$$\mathbb{E}S_n^{2m}\left(\frac{1}{2}\right) = 2^{-2m-n+1}\sum_{k=0}^{\ell}\binom{n}{k}(2k-n)^{2m}.$$

Then, (2.3) follows, since

$$\begin{aligned}&\mathbb{E}S_n^{2m+2\ell+2}\left(\frac{1}{2}\right) - \sum_{j=0}^{\ell} c_j 2^{2j-2\ell-2}\mathbb{E}S_n^{2m+2j}\left(\frac{1}{2}\right) \\ &= 2^{-2m-2\ell-n-1}\sum_{k=0}^{\ell}\binom{n}{k}a_k^{m+\ell+1}\end{aligned}$$

$$
- \sum_{j=0}^{\ell} c_j 2^{-2m-2\ell-n-1} \sum_{k=0}^{\ell} \binom{n}{k} a_k^{m+j}
$$
$$
= 2^{-2m-2\ell-n-1} \sum_{k=0}^{\ell} \binom{n}{k} a_k^m \left(a_k^{\ell+1} - \sum_{j=0}^{\ell} c_j a_k^j \right) = 0. \qquad \square
$$

For an application to the statistical estimate, we may combine Theorem 1.1 with Chebyshev's inequality to obtain the following.

Corollary 2.4. *For $\varepsilon > 0$, we have that*

$$
P\left(\left| \frac{1}{n} S_n(p) \right| > \varepsilon \right) \le \min_{1 \le m \le m_n} \left(\frac{\sqrt[2m]{\mathbb{E} S_n^{2m}(1/2)}}{n\varepsilon} \right)^{2m}.
$$

Noting the scaling invariance,

$$
\operatorname*{argmax}_{0 \le p \le 1} \mathbb{E} S_n^{2m}(p) = \operatorname*{argmax}_{0 \le p \le 1} \mathbb{E} \frac{S_n^{2m}(p)}{n^m},
$$

and $\mathbb{E} Z^{2m} = 2^{-m}(2m)!/m!$ for the standard normal random variable Z. In the limit $n \to \infty$, $\varepsilon \to 0$, $n\varepsilon^2 \to \widetilde{n}$, we have

$$
B_m := \mathbb{E} \frac{S_n^{2m}(1/2)}{n^{2m}\varepsilon^{2m}} = \mathbb{E} \frac{(S_n(1/2)/\sqrt{n/4})^{2m}}{n^{2m}\varepsilon^{2m}} \left(\frac{n}{4} \right)^m \longrightarrow 2^{-2m} \widetilde{n}^{-m} \mathbb{E} Z^{2m}
$$
$$
= 2^{-3m} \frac{(2m)!}{m!} \widetilde{n}^{-m}.
$$

In particular, we ask for the best choice of m for large n, i.e., in the above limit as $n \to \infty$, $\varepsilon \downarrow 0$, $n\varepsilon^2 \to \widetilde{n}$. The quantity $\widetilde{n} = n\varepsilon^2$ denotes an *effective sample size* in the sense of the risk assessment defined by $P(|S_n(p)| > n\varepsilon) < \varepsilon$, see **[3]** for an introduction of this artful terminology in a much broader context. Observe that, in the limit of large n,

$$
\lim_{\substack{n \to \infty \\ \varepsilon \downarrow 0 \\ n\varepsilon^2 = \widetilde{n}}} \frac{B_{m+1}}{B_m} = \frac{2m+1}{4\widetilde{n}} \begin{cases} \le 1, \\ = 1, \\ \ge 1, \end{cases}
$$

if and only if

$$m \begin{cases} \leq 2\widetilde{n} - 1/2, \\ = 2\widetilde{n} - 1/2, \\ \geq 2\widetilde{n} - 1/2. \end{cases}$$

The conclusion is perhaps best summarized in terms of the following, informally interpreted optimal estimation principle.

Approximate rule of thumb. For large n, the optimal moment order $2m$ for the Chebyshev bound is quadruple the effective sample size. In particular, the fourth moment is optimal for a one unit effective sample size.

REFERENCES

1. R.N. Bhattacharya, personal communication, 2016.

2. R.N. Bhattacharya and E.C. Waymire, *A basic course in probability theory*, Universitext, Springer, New York, 2016.

3. J. Duchi, M.J. Wainwright and M.I. Jordan, *Local privacy and minimax bounds*: *Sharp rates for probability estimation*, in *Advances in neural information processing systems*, (2013), 1529–1537.

4. D. Skinner, *Concentration of measure inequalities*, Master of Science thesis, Oregon State University, Corvallis, 2017.

University of Cologne, Mathematical Institute, Weyertal 86-90, 50931 Cologne, Germany
Email address: **cjenning@math.uni-koeln.de**

Oregon State University, Department of Mathematics, Corvallis, OR 97331
Email address: **Dane.R.Skinner@gmail.com**

Oregon State University, Department of Mathematics, Corvallis, OR 97331
Email address: **waymire@math.oregonstate.edu**