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, there it is readily observed that $\text{argmax}_{0 \le p \le 1} \mathbb{E}S_n^2(p) = \text{argmax}_{0 \le p \le 1} np(1-p) = \frac{1}{2}$, and $\mathbb{E}S_n^2(\frac{1}{2}) = \frac{n}{4}$. Rabi Bhattacharya observed that while the second moment Chebyshev sample size for a 95% confidence estimate within ± 5 percentage points is $n = 2000$, the fourth moment yields the substantially reduced polling requirement of $n = 775$. Why stop at fourth moment? Is the argmax achieved at $p = \frac{1}{2}$ for higher order moments and, if so, does it help, and compute $\mathbb{E}S_n^{2m}(\frac{1}{2})$? As captured by the title of this note, answers to these questions lead to a simple rule of thumb for 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 i.i.d. Bernoulli $0-1$ random variables $X_1, X_2, \ldots, X_n$ with parameter $p \in [0,1]$, and let $S_n(p) = \sum_{j=1}^{n}(X_j - p)$. There it is readily observed that $\text{argmax}_{0 \le p \le 1} \mathbb{E}S_n^2(p) = \text{argmax}_{0 \le p \le 1} np(1-p) = \frac{1}{2}$. It is also a well-known probability exercise to check that 4-th 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 (Bhattacharya and Waymire, 2016; p.100). Somewhat relatedly, Rabi Bhattacharya (personal communication) recently noticed, after a mildly tedious calculation to check $\text{argmax}_{0 \le p \le 1} \mathbb{E}S_n^4(p) = \frac{1}{2}$, 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% 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 (Bhattacharya and Waymire, 2016) for comparison of fourth order Chebyshev to other

Mathematical Institute, University of Cologne, 50931 Köln, Germany. jennichr@math.oregonstate.edu.
Department of Mathematics, Oregon State University, Corvallis, OR 97331. skinner@onid.oregonstate.edu.
Department of Mathematics, Oregon State University, Corvallis, OR 97331. waymire@math.oregonstate.edu.

concentration inequality bounds, and (Skinner, 2017) for numerical comparisons to higher order Chebyshev bounds.

So why stop at fourth moments? Is $\text{argmax}_{0\leq p\leq 1}\mathbb{E}S_n^{2m}(p) = \frac{1}{2}$ for all m, n and, if so, does it improve the estimate? Somewhat surprisingly we were not able 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 theorem below, part (b) provides a direct computation of $\mathbb{E}S_n^{2m}(\frac{1}{2})$. Part (c) then provides a more readily computable version.

Theorem 1.1. (a) *For all* $m \geq 1$ *and all* n *sufficiently large,* $\text{argmax}_{0\leq p\leq 1}\mathbb{E}S_n^{2m}(p) = \frac{1}{2}$.

(b) *For all positive* m *and* n, $\mathbb{E}S_n^{2m}(\frac{1}{2}) = 4^{-m}\sum_{\mu\in\pi(m),|\mu|\leq m\wedge n}\binom{2m}{2\mu_1,\ldots,2\mu_{|\mu|}}\binom{n}{|\mu|}$,

(c) *For all positive* m *and* n, $\mathbb{E}S_n^{2m}(\frac{1}{2}) = 2^{-2m-n}\sum_{k=0}^{n}\binom{n}{k}(2k-n)^{2m}$.

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,

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

The example below illustrates the challenge to locating absolute maxima for such polynomials (in p), especially to 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 simple to check using the 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 let us first observe that part (a) of the theorem does not hold for $m > n$.

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

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

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

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

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

2. Proofs and Remarks

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

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

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

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

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). That $p = \frac{1}{2}$ is a critical point follows from (*d*) of Lemma 2.1 together with (2.1) and (2.2) by examining the terms $f_i'(\mu, \frac{1}{2}) \prod_{j \ne i}^{|\mu|} f_j(\mu, \frac{1}{2})$. 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. To see that $p = \frac{1}{2}$ is an absolute maximum, the trick is to observe that for $0 \le p < \frac{1}{2} < q$, the leading coefficient of $\frac{d}{dp}\mathbb{E}S_n^{2m}(p)$, 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 $p < \frac{1}{2}$. 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 (*b*) of Lemma 2.1, it follows that $p = \frac{1}{2}$ is the unique global maximum.

For part (b) of the theorem one simply computes from independence, writing $\tilde{X}_i = X_i - \frac{1}{2}, i = 1, 2, \ldots, n$. In particular, $\tilde{X}_i = \pm\frac{1}{2}$ with equal probabilities. So, for $m \geq 1$,

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

Here one adopts 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. So nonzero contributions to the sum are provided by ordered partitions μ of size $|\mu| \leq m \wedge n$.

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

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

□

Remark 2.1. *A linear recurrence in m is possible to aid the pre-asymptotic (in n) computation of $\mathbb{E}S_n^{2m}(\frac{1}{2})$. Namely,*

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

where $\ell = \left\lfloor \frac{n-1}{2} \right\rfloor$, $a_k = (2k-n)^2$, and $(c_0, c_1, \ldots, c_\ell)$ is the unique solution to

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

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*

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

For the application to statistical estimation one may combine Theorem 1.1 with Chebyshev's inequality to obtain,

Corollary 2.1. *For $\epsilon > 0$, we have that $P\left(|\frac{1}{n}S_n(p)| > \epsilon\right) \leq \min_{1\leq m\leq m_n}\left(\frac{\sqrt[2m]{\mathbb{E}S_n^{2m}(\frac{1}{2})}}{n\epsilon}\right)^{2m}$.*

Noting the scaling invariance $\mathrm{argmax}_{0\leq p\leq 1}\mathbb{E}S_n^{2m}(p) = \mathrm{argmax}_{0\leq p\leq 1}\mathbb{E}\frac{S_n^{2m}(p)}{n^m}$, and $\mathbb{E}Z^{2m} = 2^{-m}\frac{(2m)!}{m!}$ for the standard normal random variable Z, in the limit "$n\to\infty, \epsilon\to 0, n\epsilon^2\to\tilde{n}$" one has

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

In particular, one may ask for the best choice of m for large n, i.e, in the above limit as $n\to\infty, \epsilon\downarrow 0, n\epsilon^2\to\tilde{n}$. The quantity $\tilde{n} = n\epsilon^2$ denotes an *effective sample size* in the sense of the risk assessment defined by $P(|S_n(p)| > n\epsilon) < \epsilon$; see (Duchi et al. 2013) for an introduction of this artful terminology in a much broader context. Observe that in the limit of large n

$$\lim_{n\to\infty,\epsilon\downarrow 0, n\epsilon^2=\tilde{n}}\frac{B_{m+1}}{B_m} = \frac{2m+1}{4\tilde{n}}\begin{cases}\leq 1\\ = 1\\ \geq 1\end{cases}$$

if and only if

$$m\begin{cases}\leq 2\tilde{n}-\frac{1}{2}\\ = 2\tilde{n}-\frac{1}{2}\\ \geq 2\tilde{n}-\frac{1}{2}.\end{cases}$$

The take-away 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!*

3. References

Bhattacharya, R.N., and Waymire, E.C. (2016), "A Basic Course in Probability Theory", 2nd ed., Universitext, Springer, NY.

Duchi, J. and Wainwright, M. J. and Jordan, M. I. (2013), "Local privacy and minimax bounds: Sharp rates for probability estimation", in Advances in Neural Information Processing Systems, 1529-1537.

Skinner, Dane, 2017, "Concentration of Measure Inequalities", Master of Science, Oregon State University