Combinatorics, Probability and Computing (1997) 6, 79–86
Copyright © 1997 Cambridge University Press

Centering Sequences with Bounded Differences

COLIN McDIARMID

Department of Statistics, University of Oxford, Oxford, UK

Received 17 September 1993; revised 7 April 1995

Inequalities for martingales with bounded differences have recently proved to be very useful in combinatorics and in the mathematics of operational research and computer science. We see here that these inequalities extend in a natural way to 'centering sequences' with bounded differences, and thus include, for example, better inequalities for sequences related to sampling without replacement.

1. Centering sequences

Given a sequence $\mathbf{X} = (X_1, X_2, \ldots)$ of (integrable) random variables the corresponding difference sequence is $\mathbf{Y} = (Y_1, Y_2, \ldots)$, where $Y_k = X_k - X_{k-1}$ (and where we always set $X_0 \equiv 0$). Let $\mu_k(x) = E(Y_k | X_{k-1} = x)$. More formally, let the function $\mu_k(x)$ be such that $\mu_k(X_{k-1})$ is a version of $E(Y_k | X_{k-1})$. Of course, $\mu_k(x)$ is not uniquely defined for all x, but this will not cause any difficulties. We call the distribution of the sequence $\mathbf{X}$ *centering* if for each $k = 2, 3, \ldots$ we may take $\mu_k(x)$ to be a non-increasing function of x.

A centering distribution is 'self-balancing' or 'self-correcting'. If X_k is an individual's capital at time k, then having a centering distribution corresponds to the worthy maxim 'the less you have the more you get on average'. Observe that a martingale is trivially centering, since we may take each $\mu_k(x) \equiv 0$. Also, various time series (economic or other) that are controlled towards some target should be centering. Further, note that if c and $a_1, a_2, a_3, \ldots$ are constants and $X_1, X_2, X_3 \ldots$ is a centering sequence then so is $cX_1 + a_1, cX_2 + a_2, cX_3 + a_3, \ldots$.

Inequalities for martingales with bounded differences due to Hoeffding [10] and others and presented in [5, 18] have proved to be very useful in combinatorics and in the mathematics of operational research and theoretical computer science. We shall see here that these inequalities extend naturally and easily to centering sequences with bounded differences. The most basic example for the martingale inequalities involves the binomial distribution: here we include the hypergeometric distribution naturally in the same inequalities. Before we consider the inequalities let us discuss four examples of centering sequences.

https://doi.org/10.1017/S0963548396002854 Published online by Cambridge University Press

Example 1: Sampling without replacement

Let $1 \leq n \leq N$ and let $(x_1, \ldots, x_N) \in \mathbb{R}^N$. Choose a random sample S of size n without replacement from $\{1, \ldots, N\}$ and let $X = \sum_{j \in S} x_j$. Note that when each $x_j = 0$ or 1 then the random variable X has the hypergeometric distribution. Now $EX = nr/N$ where $r = \sum_{j=1}^{N} x_j$. We are interested in bounding quantities like $P(|X - EX| \geq t)$.

Let $(Z_1, \ldots, Z_n)$ be a random n-tuple of *distinct* elements of $\{1, \ldots, N\}$, all such n-tuples being equally likely. Let $Y_j = x_{Z_j}$ and $X_k = \sum_{j=1}^{k} Y_j$. Of course, X_n has the same distribution as X. The interesting fact for us here is that the sequence $\mathbf{X} = (X_1, \ldots, X_n)$ is easily seen to be centering, since

$$\mu_k(x) = E(Y_k | X_{k-1} = x) = \frac{r - x}{N - k + 1},$$

which is a decreasing function of x.

When the results in the next section are applied to this problem of sampling without replacement we obtain inequalities of Hoeffding (see Theorem 4 of [10]). Let us restrict our attention now to the special case of the hypergeometric distribution (when each $x_j = 0$ or 1). For a neat proof of such an inequality see [6]. From Theorem 2.2(c) below, it will follow for example that for any $\varepsilon > 0$,

$$P(X \leq (1 - \varepsilon)EX) < \exp(-\frac{1}{2}\varepsilon^2 EX).$$

If we try to apply here the usual inequalities for martingales with bounded differences [18] in the natural way we obtain an unwanted factor < 1 in the exponent in the bound.

Example 2: Occupancy problems

Suppose that we have N boxes, where box j has weight w_j, and we throw balls one after another uniformly at random into these boxes. Let X_k be the total weight of the empty boxes after k balls have been thrown. Then the sequence $X_1, X_2, \ldots$ is centering, since

$$\mu_k(x) = E(X_k - X_{k-1} | X_{k-1} = x) = -\frac{x}{N}.$$

The most interesting case is when each $w_j = 1$. Theorem 2.2(b) below includes 'occupancy bound 2' (Theorem 3) of [12]. (In occupancy bound 1 of that paper, we gain a factor 4 in the exponent of the bound if we use the bounded differences inequality Corollary 6.10 of [18] rather than Azuma's inequality.)

Example 3: Binary search trees

Consider a binary search tree formed by repeatedly inserting random keys, where the keys are distinct and all linear orders on the keys are equally likely (see, for example, [14, 17, 19]). Thus, after inserting two keys k, k' there is a root containing k, and a son (a leaf) containing k' which is a leftson if $k' < k$ and a rightson if $k < k'$. Let L_n be the number of leaves in the tree when it has n nodes, so that $L_1 = L_2 = 1$. Then $\mathbf{L} = (L_1, L_2, \ldots)$ is a Markov chain; and conditional on $L_{k-1} = x$ we have $L_k = x$ with probability $\frac{2x}{k}$ and $L_k = x + 1$ otherwise. Thus the sequence $\mathbf{L}$ is centering, since

$$\mu_k(x) = E(L_k - L_{k-1} | L_{k-1} = x) = 1 - \frac{2x}{k},$$

which is a decreasing function of x.

https://doi.org/10.1017/S0963548396002854 Published online by Cambridge University Press

We also have that $EL_k = 1 + (1 - \frac{2}{k})EL_{k-1}$ and so $EL_k = \frac{k+1}{3}$ for $k \geq 2$ [16]. From Theorems 2.2 or 2.3 below, it will follow that

$$P\left(\left|L_n - \frac{n+1}{3}\right| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{n}\right)$$

for any $n \geq 2$ and $t > 0$.

Now let $N_k^{(i)}$ denote the number of nodes with exactly i children when the tree has n nodes, so that $N_k^{(0)} = L_k$. Then $N_k^{(1)} = k + 1 - 2L_k$ and $N_k^{(2)} = L_k - 1$. Hence, for each $i = 0, 1, 2$ the sequence $N_1^{(i)}, N_2^{(i)}, N_3^{(i)}, \ldots$ is centering. Similar results hold for the 'fringe analysis' of search trees, see for example [1] and the references in that paper.

Example 4: Urn models

An urn contains red and black balls. A ball is drawn at random from the urn and removed. If the ball was red then $n(r,r)$ red balls and $n(r,b)$ black balls are added back into the urn; and if the ball was black then $n(b,r)$ red balls and $n(b,b)$ black balls are added. We shall assume that the numbers $n(x,y)$ are each non-negative integers, and

$$n(r,r) + n(r,b) = n(b,r) + n(b,b) = c.$$

If we start with N balls then after n rounds there is a total of $N + n(c-1)$ balls in the urn. Let R_n be the number of red balls after n rounds. It is easily checked that the sequence $R_1, R_2, \ldots$ is centering if and only if $n(r,r) \leq n(b,r) + 1$. In the language of Feller (see [8] V.2), we are interested not in 'contagion' (where we tend to see runs of similar events) but in 'safety campaigns' (where we tend to see alternations).

The case when we start with r red balls and $N - r$ black balls, and all the $n(x,y)$ are 0, corresponds to sampling without replacement, as considered in the first example (with each $x_j = 0$ or 1). The case when we start with N red balls and no black ones, and let $n(r,r) = n(b,r) = 0$ and $n(b,b) = n(r,b) = 1$ (that is, we always add back exactly one black ball), corresponds to the occupancy problem considered in Example 2 (with each $x_j = 1$).

Finally, consider the third example. Suppose that we start with two red balls and no black balls. (Red balls correspond to external nodes that are sons of leaves and black balls to other external nodes.) Let $n(r,r) = n(r,b) = 1$, $n(b,r) = 2$ and $n(b,b) = 0$. Then $R_1, R_2, R_3, \ldots$ has the same distribution as $2L_1, 2L_2, 2L_3, \ldots$.

2. Inequalities for centering sequences with bounded differences

Theorem 2.1. *Let $X_1, X_2, \ldots, X_n$ be a centering sequence with corresponding differences $Y_k = X_k - X_{k-1}$ satisfying $0 \leq Y_k \leq 1$ for each k. Let $\alpha = \frac{1}{n}E(X_n)$, and let $\bar{\alpha} = 1 - \alpha$. Then for any $0 \leq t < \bar{\alpha}$,*

$$P(X_n - EX_n \geq nt) \leq \left(\left(\frac{\alpha}{\alpha + t}\right)^{\alpha+t}\left(\frac{\bar{\alpha}}{\bar{\alpha} - t}\right)^{\bar{\alpha}-t}\right)^n.$$

This result will be proved in the next section. Note that it yields immediately the fact that the same bound holds for $P(X_n - EX_n \leq nt)$ if $-\alpha < t \leq 0$ (just consider the centering sequence $k - X_k$). The following corollary is perhaps more useful.

https://doi.org/10.1017/S0963548396002854 Published online by Cambridge University Press

Theorem 2.2. *As in Theorem* 2.1 *let* $X_1, X_2, \ldots, X_n$ *be a centering sequence with corresponding differences* $Y_k = X_k - X_{k-1}$ *satisfying* $0 \le Y_k \le 1$ *for each* k.
(a) *For any* $t > 0$,

$$P(X_n \ge EX_n + t) \le \exp(-2t^2/n),$$
$$P(X_n \le EX_n - t) \le \exp(-2t^2/n).$$

(b) *For any* $\varepsilon > 0$,

$$\begin{aligned} P(X_n \ge (1+\varepsilon)EX_n) &\le \exp(-((1+\varepsilon)\log(1+\varepsilon) - \varepsilon)EX_n) \\ &\le \exp(-(\varepsilon^2/2 - \varepsilon^3/6)EX_n) \\ &\le \exp(-(\varepsilon^2/3)EX_n) \text{ if also } \varepsilon \le 1. \end{aligned}$$

(c) *For any* $\varepsilon > 0$,

$$P(X_n \le (1-\varepsilon)EX_n) \le \exp\left(-\frac{1}{2}\varepsilon^2 EX_n\right).$$

The above results really refer to the case when we have a uniform bound on the range of the differences Y_k, that is we have constants $c > 0$ and $a_1, \ldots, a_n$ such that $a_k \le Y_k \le a_k + c$ for each k. For in this case, we may let $\hat{Y}_k = (Y_k - a_k)/c$ so that $0 \le \hat{Y}_k \le 1$ and let $\hat{X}_k = \sum_{j=1}^{k} \hat{Y}_j$. Then $\hat{X}_1, \ldots, \hat{X}_n$ is centering and $X_n - EX_n = c(\hat{X}_n - E\hat{X}_n)$.

Our third and final theorem extends part (a) of Theorem 2.2 and extends Corollary (6.9) of [18].

Theorem 2.3. *Let* $X_1, X_2, \ldots, X_n$ *be a centering sequence with corresponding differences* $Y_k = X_k - X_{k-1}$, *and suppose that there are constants* a_k, b_k *such that* $a_k \le Y_k \le b_k$ *for each* k. *Then for any* $t > 0$,

$$P(X_n - EX_n \ge t) \le \exp\left(-2t^2 / \sum (b_k - a_k)^2\right),$$
$$P(X_n - EX_n \le -t) \le \exp\left(-2t^2 / \sum (b_k - a_k)^2\right).$$

3. Proofs

We shall use the following well known inequality of Chebyshev (see [9] p. 43).

Lemma 3.1. *Let* X *be a real-valued random variable, let* $f : \mathbb{R} \to \mathbb{R}$ *be a non-decreasing function and let* $g : \mathbb{R} \to \mathbb{R}$ *be a non-increasing function. Then*

$$E(f(X)g(X)) \le E(f(X))E(g(X)).$$

(We assume here and elsewhere that all the functions involved are measurable and that all expectations exist.)

Proof. Let Y be an independent copy of X. Note that

$$(f(X) - f(Y))(g(X) - g(Y)) \le 0$$

almost surely, and take expectations. □

https://doi.org/10.1017/S0963548396002854 Published online by Cambridge University Press

We need one more preliminary lemma which is taken from [10] (and which is also contained in the proof of Lemma (5.8) of [18]).

Lemma 3.2. *Let X be a real-valued random variable with $EX = \mu$ and $a \le X \le b$ where a and b are constants. Then for any h*

$$\begin{aligned} E\left(e^{hX}\right) &\le \left(\frac{b-\mu}{b-a}e^{ha} + \frac{\mu-a}{b-a}e^{hb}\right) \\ &\le e^{h\mu}\exp\left(\frac{1}{8}h^2(b-a)^2\right). \end{aligned}$$

□

Now we reach the heart of the matter.

Lemma 3.3. *Let $X_1, X_2, \ldots, X_n$ be a centering sequence with corresponding differences $Y_k = X_k - X_{k-1}$ with mean $\mu_k = EY_k$, and suppose that there are constants a_k, b_k such that $a_k \le Y_k \le b_k$ for each k. Then for any $h > 0$*

$$E(\exp hX_n) \le \prod_{k=1}^{n}\left(\frac{b_k-\mu_k}{b_k-a_k}e^{ha_k} + \frac{\mu_k-a_k}{b_k-a_k}e^{hb_k}\right).$$

Proof. We may assume that $\mu_n(x)$ is a non-increasing function of x, where $\mu_n(X_{n-1}) = E(Y_n|X_{n-1})$. For any h,

$$\begin{aligned} E(\exp hX_n) &= E\left(\exp(hX_{n-1})E\left(\exp(hY_n)\mid X_{n-1}\right)\right) \\ &\le E\left(\exp(hX_{n-1})\left(\frac{b_n-\mu_n(X_{n-1})}{b_n-a_n}e^{ha_n} + \frac{\mu_n(X_{n-1})-a_n}{b_n-a_n}e^{hb_n}\right)\right), \end{aligned}$$

by Lemma 3.2. But for $h > 0$

$$\frac{b-t}{b-a}e^{ha} + \frac{t-a}{b-a}e^{hb}$$

is an increasing function of t, and so the second factor in the expectation above is a non-increasing function of X_{n-1}. Hence by Lemma 3.1, for $h > 0$

$$E(\exp hX_n) \le E(\exp hX_{n-1})\left(\frac{b_n-\mu_n}{b_n-a_n}e^{ha_n} + \frac{\mu_n-a_n}{b_n-a_n}e^{hb_n}\right)$$

and we may complete the proof by induction on n. □

Proof of Theorem 2.1. We apply Lemma 3.3 with each $a_k = 0$ and $b_k = 1$ and obtain:

$$\begin{aligned} E(\exp hX_n) &\le \prod_{k=1}^{n}\left((1-\mu_k) + \mu_k e^h\right) \\ &\le \left(1-\alpha+\alpha e^h\right)^n, \end{aligned}$$

using the fact that the geometric mean is at most the arithmetic mean. Hence, for any $h > 0$

$$\begin{aligned} P(X_n \ge n(\alpha+t)) &\le \exp(-hn(\alpha+t))E(\exp hX_n) \\ &\le \left(e^{-h(\alpha+t)}(1-\alpha+\alpha e^h)\right)^n. \end{aligned}$$

https://doi.org/10.1017/S0963548396002854 Published online by Cambridge University Press

We now obtain the desired bound by setting

$$e^h = \frac{(\alpha + t)\bar{\alpha}}{\alpha(\bar{\alpha} - t)}. \qquad \square$$

Proof of Theorem 2.2. Parts (a) and (c) follow from Theorem 2.1, exactly as in [18] the corresponding parts of Corollary (5.2) follow from Theorem (5.1). To prove part (b) it is simpler to use the proof of Theorem 2.1. If we set $t = \varepsilon\alpha$ and $e^h = (1 + \varepsilon)$ there, and use the inequality $1 + x \leq e^x$, we obtain the first inequality in (b) (see also Appendix A of [2]). Now let $f(x) = ((1 + x)\log(1 + x) - x) - (x^2/2 - x^3/6)$ for $x \geq 0$. Then $f(0) = 0$, and $f'(x) = \log(1 + x) - (x - x^2/2) > 0$ for $x > 0$. Hence $f(x) > 0$ for $x > 0$. The second inequality in (b) now follows, and the third is trivial. $\square$

Proof of Theorem 2.3. Our proof here follows the lines of the proof of Theorem (6.7) in [18]. By Lemmas 3.2 and 3.3, for $h > 0$

$$\begin{aligned} E(\exp hX_n) &\leq \prod_{k=1}^{n} \exp\left(hEY_k + \frac{1}{8}h^2(b_k - a_k)^2\right) \\ &= \exp(hEX_n)\exp\left(\frac{1}{8}h^2\sum(b_k - a_k)^2\right). \end{aligned}$$

Hence for $h > 0$

$$\begin{aligned} P(X_n - EX_n \geq t) &\leq e^{-ht}E(\exp h(X_n - EX_n)) \\ &\leq e^{-ht}\exp\left(\frac{1}{8}h^2\sum(b_k - a_k)^2\right). \end{aligned}$$

Now set $h = 4t/\sum(b_k - a_k)^2$ to obtain the former inequality in the theorem. To deduce the latter replace $\mathbf{X}$ by $-\mathbf{X}$. $\square$

4. Concluding remarks

4.1. Negative dependence

For a sequence of just two random variables X_1, X_2 our definition of centering is a weakened form of Lehman's [15] definition of *negative regression dependence*, and in general the definition is like a weakened form of being *conditionally decreasing in sequence* [7]. For related definitions see, for example, [3, 4, 7, 11, 15].

Recall that two random variables X and Y are said to be *negative quadrant (or orthant) dependent* if $P(X > x, Y > y) \leq P(X > x)P(Y > y)$ for all x, y. If we have a sequence $\mathbf{X} = (X_1, X_2, \ldots)$ of random variables such that for each $k = 2, 3, \ldots$ the random variables X_{k-1} and $Y_k = X_k - X_{k-1}$ are negative quadrant dependent, then it follows that

$$E(e^{hX_n}) \leq \prod_{k=1}^{n} E(e^{hY_k})$$

for any h. It follows that all our bounds for centering sequences hold also for such a sequence $\mathbf{X}$.

https://doi.org/10.1017/S0963548396002854 Published online by Cambridge University Press

4.2. Bounds on maxima

All the inequalities given in [18] for independent random variables and more generally for martingale difference sequences, and all the inequalities given here for centering sequences, are based on the Bernstein inequality, that $P(X \geq t) \leq e^{-ht}E(e^{hX})$ for any $h > 0$. In Section 6(c) of [18], it is noted that, when we start with a martingale difference sequence, we may thus use Doob's maximal inequality for submartingales to strengthen some of the inequalities, by replacing $P(X_n - EX_n \geq t)$ by $P(\max_k(X_k - EX_k) \geq t)$. Here we are unable to obtain such strengthenings.

4.3. Variances

Let $X_1, \ldots, X_n$ be a centering sequence with differences Y_k. By Lemma 3.1, the covariance of X_{k-1} and Y_k satisfies

$$cov(X_{k-1}, Y_k) = E\left((X_{k-1} - EX_{k-1})E(Y_k|X_{k-1})\right) \leq 0,$$

and so

$$var(X_n) \leq \sum_{k=1}^{n} var(Y_k).$$

Assume now that $a_k \leq Y_k \leq b_k$ for constants a_k, b_k as in Theorem 2.3. Then $var(Y_k) \leq \frac{1}{4}(b_k - a_k)^2$, and so $var(X_n) \leq \frac{1}{4}s^2$ where $s^2 = \sum(b_k - a_k)^2$.

Theorem 2.3 also yields a bound on $var(X_n)$ but we lose the factor $\frac{1}{4}$. For

$$\begin{aligned} var(X_n) &= \int_0^\infty P((X_n - EX_n)^2 \geq t)dt \\ &\leq \int_0^\infty 2\exp\left(-\frac{2t}{s^2}\right)dt \\ &= s^2. \end{aligned}$$

Using $var(X_n) \leq \frac{1}{4}s^2$ together with Chebyshev's inequality gives better bounds on $P(|X_n - EX_n| \geq t)$ for small t than does Theorem 2.3 above, though of course the theorem is much better for large t.

Acknowledgements

I would like to thank David Aldous, David Stirzaker and a referee for helpful comments.

References

[1] Aldous, D., Flannery, B. and Palacios, J. L. (1988) Two applications of urn processes. *Probability in the Engineering and Informational Sciences* **2** 293–307.
[2] Alon, N. and Spencer, J. (1992) *The Probabilistic Method*. Wiley.
[3] Block, H. W., Savits, T. H. and Shaked, M. (1982) Some concepts of negative dependence. *Annals of Probability* **10** 765–772.
[4] Block, H. W., Savits, T. H. and Shaked, M. (1989) A concept of negative dependence using stochastic ordering. *Statistics and Probability Letters* **3** 81–86.
[5] Bollobás, B. (1987) Martingales, isoperimetric inequalities and random graphs. *Colloq. Math. Soc. János Bolyai* **52** 113–139.

https://doi.org/10.1017/S0963548396002854 Published online by Cambridge University Press

[6] Chvátal, V. (1979) The tail of the hypergeometric distribution. *Discrete Mathematics* **25** 285–287.
[7] Ebrahimi, N. and Ghosh, M. (1981) Multivariate negative dependence. *Communications in Statistics A: Theory and Methods* **10** 307–337.
[8] Feller, W. J. (1968) *An Introduction to Probability Theory and its Applications*, Vol 1, 3rd ed. Wiley.
[9] Hardy, G., Littlewood, J. E. and Pólya, G. (1952) *Inequalities* 2nd ed. Cambridge University Press.
[10] Hoeffding, W. J. (1963) Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 713–721.
[11] Joag-Dev, K. and Proschan, F. (1983) Negative association of random variables with applications. *Annals of Statistics* **11** 286–295.
[12] Kamath, A., Motwani, R., Palem, K. and Spirakis, P. (1995) Tail bounds for occupancy and the satisfiability threshold conjecture. *Random Structures and Algorithms* **7** 59–80.
[13] Karlin, S. and Rinott, Y. (1980) Classes of orderings of measures and related correlation inequalities II: multivariate reverse rule distributions. *J. Multivariate Anal.* **10** 499–516.
[14] Knuth, D. E. (1973) *The Art of Computer Programming: Vol 3, Sorting and Searching*. Addison-Wesley.
[15] Lehman, E. L. (1966) Some concepts of dependence. *Ann. Math. Statistics* **37** 1137–1153.
[16] Mahmoud, H. M. (1986) The expected distribution of degrees in random binary search trees. *The Computer Journal* **29** 36–37.
[17] Mahmoud, H. M. (1992) *Evolution of Random Search Trees*. Wiley.
[18] McDiarmid, C. (1989) On the method of bounded differences. In J. Siemons (ed.), *Surveys in Combinatorics: London Math. Soc. Lecture Note Series 141*. Cambridge University Press.
[19] Sedgewick, R. (1988) *Algorithms* 2nd ed. Addison-Wesley.

https://doi.org/10.1017/S0963548396002854 Published online by Cambridge University Press