Some new maximal inequalities

Peter Harremoës *

Centrum voor Wiskunde en Informatica, PO Box 94079, 1098 SJ Amsterdam, The Netherlands

ABSTRACT

New maximal inequalities for non-negative martingales are proved. The inequalities are tight and strengthen well-known maximal inequalities by Doob. The inequalities relate martingales to information divergence and imply convergence of $X \ln X$ bounded martingales. Similar results hold for stationary sequences.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Comparing results from probability theory and information theory is not a new idea. Many convergence theorems in probability theory can be reformulated as "the entropy converges to its maximum" or "information divergence converges to zero ". The Weak Law of Large Numbers as well as its generalizations, mean convergence of martingales and stationary sequences, can be proved using information theoretic techniques, see Csiszár (1963), Barron (1991) and Csiszár and Shields (2004). Large deviation bounds are closely related to information theory and can be used to prove the Strong Law of Large Numbers (Csiszár and Shields, 2004, p. 13). In this paper we shall see that information theoretic ideas are also relevant for pointwise convergence of martingales and stationary sequences. In this short paper the focus is on martingales.

Let $(\Omega, \mathbb{B}, Q)$ be a probability space. If not otherwise stated $E[\cdot]$ shall denote mean value with respect to Q. The following inequalities are well known and a proof can be found in Shiryaev (1996, p. 494).

Lemma 1 (*Doob's Maximal Inequalities*). *Let* $(X_1, \mathcal{F}_1), \ldots, (X_n, \mathcal{F}_n)$ *be a non-negative martingale. Let the random variables* $X^{\max}$ *and* $X^{\min}$ *be given by* $X^{\max} = \max\{X_j\}$ *and* $X^{\min} = \min\{X_j\}$. *Then the following inequalities hold*

$$\lambda \cdot Q\left(X^{\max} \geq \lambda\right) \leq E\left[X_n \cdot 1_{X^{\max} \geq \lambda}\right] \tag{1}$$

$$\lambda \cdot Q\left(X^{\min} \leq \lambda\right) \geq E\left[X_n \cdot 1_{X^{\min} \leq \lambda}\right]. \tag{2}$$

A similar inequality holds for ergodic sequences, see Shiryaev (1996, p. 410).

Lemma 2 (*Maximal Ergodic Theorem*). *Let* $T : \Omega \to \Omega$ *denote a measurable transformation that conserves the probability measure* Q. *Let* f *be a random variable with* $E|f| < \infty$. *Define* $f_n^{\min}$ *and* $f_n^{\max}$ *by*

$$f_n^{\min} = \min_{1 \leq k \leq n} \frac{1}{k} \sum_{j=0}^{k-1} f \circ T^j \tag{3}$$

$$f_n^{\max} = \max_{1 \leq k \leq n} \frac{1}{k} \sum_{j=0}^{k-1} f \circ T^j. \tag{4}$$

---

* Tel.: +31 0 20 592 4272; fax: +31 0 20 592 4199.
  E-mail address: P.Harremoes@cwi.nl.

Then

$$\lambda \cdot Q \left(f_n^{\max} \geq \lambda \right) \leq E \left[f \cdot 1_{f_n^{\max} \geq \lambda} \right] \tag{5}$$

$$\lambda \cdot Q \left(f_n^{\min} \leq \lambda \right) \geq E \left[f \cdot 1_{f_n^{\min} \leq \lambda} \right]. \tag{6}$$

2. Some new maximal inequalities

The following function will play an important role in what follows. Let $\gamma(x) = x - 1 - \ln x$ for $x > 0$. Remark that γ is strictly convex with minimum $\gamma(1) = 0$.

Theorem 3. *Let* $(X_1, \mathcal{F}_1), (X_2, \mathcal{F}_2), \ldots (X_n, \mathcal{F}_n)$ *be a non-negative martingale. Let* $X^{\max} = \max(X_j)$ *and* $X^{\min} = \min(X_j)$. *If* $X_1 = 1$ *then*

$$\gamma \left(E \left[X^{\max} \right] \right) \leq E \left[X_n \ln(X_n) \right] \tag{7}$$

and

$$\gamma \left(E \left[X^{\min} \right] \right) \leq E \left[X_n \ln(X_n) \right]. \tag{8}$$

Proof. By using that $X^{\max} \geq X_1 = 1$ we get

$$
\begin{aligned}
E[X^{\max}] - 1 &= \int_0^\infty Q(X^{\max} \geq t) \, dt - 1 = \int_1^\infty Q(X^{\max} \geq t) \, dt \\
&\leq \int_1^\infty \frac{1}{t} E[X_n \cdot 1_{X^{\max} \geq t}] \, dt = E \left[\int_1^\infty \frac{X_n \cdot 1_{X^{\max} \geq t}}{t} \, dt \right] \\
&= E \left[X_n \cdot \int_1^{X^{\max}} \frac{1}{t} \, dt \right] = E \left[X_n \cdot \ln(X^{\max}) \, dt \right].
\end{aligned}
\tag{9}
$$

Using that γ is non-negative we get

$$
\begin{aligned}
E[X^{\max}] - 1 &\leq E \left[X_n \left(\ln(X^{\max}) + \gamma \left(\frac{X^{\max}}{X_n \cdot E[X^{\max}]} \right) \right) \right] \\
&= E[X_n \ln(X_n)] + \ln E[X^{\max}].
\end{aligned}
\tag{10}
$$

Inequality (7) is obtained by reorganizing the terms.

Similarly, $0 \leq X^{\min} \leq X_1 = 1$ implies that

$$
\begin{aligned}
E \left[X^{\min} \right] &= \int_0^1 Q \left(X^{\min} \geq t \right) dt = 1 - \int_0^1 Q \left(X^{\min} < t \right) dt \\
&\leq 1 - \int_0^1 \frac{1}{t} E[X_n \cdot 1_{X^{\min} < t}] \, dt = 1 - E \left[\int_0^1 \frac{X_n \cdot 1_{X^{\min} < t}}{t} \, dt \right] \\
&= 1 - E \left[X_n \cdot \int_{X^{\min}}^1 \frac{1}{t} \, dt \right] = 1 + E \left[X_n \cdot \ln \left(X^{\min} \right) \right].
\end{aligned}
\tag{11}
$$

Inequality (8) can now be proved in the same way as (7). □

A similar inequality holds for an ergodic transformation T and a non-negative function f with $\int f \, dQ = 1$. The inequalities (7) and (8) are tight in the following sense.

Theorem 4. *Assume that* $g : \mathbb{R}_+ \to \mathbb{R}$ *is a continuous function such that*

$$g \left(E[X^{\max}] \right) \leq E[X_n \ln(X_n)] \tag{12}$$

and

$$g \left(E \left[X^{\min} \right] \right) \leq E[X_n \ln(X_n)]. \tag{13}$$

for all non-negative martingales $(X_1, \mathcal{G}_1), (X_2, \mathcal{G}_2), \ldots, (X_n, \mathcal{G}_n)$. *Then* $g \leq \gamma$.

Proof. Let g denote a function satisfying inequality (12) for all martingales. Assume that $t \in [1; \infty[$. We shall prove that $g(t) \le \gamma(t)$. For an integer k, define $n = 2^k$. For $j = 0, 1, \ldots, 2^k - 1$ let $\mathcal{F}_j^k$ denote the σ-algebra generated by the sets

$$\left[0; \frac{1}{2^k}\right[$$

$$\left[\frac{1}{2^k}; \frac{2}{2^k}\right[$$

$$\left[\frac{2}{2^k}; \frac{3}{2^k}\right[$$

$$\vdots$$

$$\left[\frac{j-1}{2^k}; \frac{j}{2^k}\right[$$

$$\left[\frac{j}{2^k}; 1\right].$$

Further $\mathcal{F}_{2^k}^k$ shall denote the Borel σ-algebra on $[0; 1]$. With these definitions $\mathcal{F}_0^k, \mathcal{F}_1^k, \ldots \mathcal{F}_{2^k}^k$ is an increasing sequence of σ-algebras. Let Q denote the uniform distribution on $[0; 1]$. Let f denote the function (random variable) on $[0; 1]$ given by $f(x) = \alpha (1 - x)^{\alpha - 1}$ where $\alpha = t^{-1} \in]0; 1]$. A martingale $\left(X_j, \mathcal{F}_j^k\right)_{j=0, 1, \ldots, 2^k}$ is given by the random variables $X_j = E\left[f | \mathcal{F}_j^k\right]$ where

$$E\left[f \,\middle|\, \mathcal{F}_j^k\right](x) = \begin{cases} \dfrac{\int_{\frac{l-1}{2^k}}^{\frac{l}{2^k}} f(y)\, \mathrm{d}y}{1/2^k} & \text{for } x \in \left[\dfrac{l-1}{2^k}; \dfrac{l}{2^k}\right[, \text{ where } l \le j, \\[3ex] \dfrac{\int_{\frac{j}{2^k}}^{1} f(y)\, \mathrm{d}y}{1 - \frac{j}{2^k}} & \text{for } x \in \left[\dfrac{j}{2^k}; 1\right]. \end{cases}$$

Remark that f is increasing and therefore

$$\max_j E\left[f \,\middle|\, \mathcal{F}_j^k\right](x) = \frac{\int_{\frac{\lfloor x2^k \rfloor}{2^k}}^{1} f(y)\, \mathrm{d}y}{1 - \frac{\lfloor x2^k \rfloor}{2^k}} = \left(1 - \frac{\lfloor x2^k \rfloor}{2^k}\right)^{\alpha - 1}.$$

For any $k \in \mathbb{N}$ we have

$$g\left(\int_0^1 \left(1 - \frac{\lfloor x2^k \rfloor}{2^k}\right)^{\alpha - 1} \mathrm{d}x\right) \le E[X_n \ln(X_n)] = \int_0^1 f(x) \log f(x)\, \mathrm{d}x$$

$$= \int_0^1 \alpha (1 - x)^{\alpha - 1} \log\left((\alpha - 1 + 1)(1 - x)^{\alpha - 1}\right) \mathrm{d}x = \log(\alpha) + \frac{1}{\alpha} - 1.$$

The sequence

$$\left(1 - \frac{\lfloor x2^k \rfloor}{2^k}\right)^{\alpha - 1}$$

is an increasing sequence of positive functions that converges to the function $(1 - x)^{\alpha}$ for k tending to infinity. According to Lebesgue's theorem the integral

$$\int_0^1 \left(1 - \frac{\lfloor x2^k \rfloor}{2^k}\right)^{\alpha - 1} \mathrm{d}x$$

converges to

$$\int_0^1 (1 - x)^{\alpha}\, \mathrm{d}x = \left[\frac{-(1 - x)^{\alpha}}{\alpha}\right]_0^1 = \frac{1}{\alpha}.$$

Thus

$$g\left(\frac{1}{\alpha}\right) \le \log(\alpha) + \frac{1}{\alpha} - 1$$

so that

$$g(t) \le -\log(t) + t - 1.$$

The proof works essentially the same way for $t \in \;]0;\,1]$ and therefore inequality (8) is also tight. □

The proof of Theorem 4 gives an indication of how Theorem 3 can be generalized to martingales with continuous time and proved via discretizations.

3. Convergence of martingales and ergodic sequences

In order to prove the convergence of martingales we have to reorganize our inequalities somewhat. Let P and Q be probability measures on the same space. Then the *information divergence from P to Q* is defined by $D(P \parallel Q) = \int \ln \frac{\mathrm{d}P}{\mathrm{d}Q}\,\mathrm{d}P$ if P is absolutely continuous with respect to Q and by $D(P \parallel Q) = \infty$ otherwise.

Theorem 5. *Let* $(X_1, \mathcal{F}_1), (X_2, \mathcal{F}_2), \ldots$ *be a non-negative martingale and assume that* $E[X_j] = 1$. *Let* P_j *be the probability measure given by* $\frac{\mathrm{d}P_j}{\mathrm{d}Q} = X_j$. *For* $m \le n$ *put* $X_{m,n}^{\max} = \sup_{j=m,\ldots,n} X_j$ *and let* $\gamma(x) = x - 1 - \ln(x)$. *Then*

$$\gamma\left(E\left[X_{m,n}^{\max}\right]\right) \le D(P_n \parallel P_m). \tag{14}$$

Proof. We note that $X_m = 0$ implies that $X_j = 0$ for $j \ge m$ and therefore also $X_{m,n}^{\max} = 0$. With the convention that $\frac{X_{m,n}^{\max}}{X_m} = 0$ when both numerator and denominator are zero we get

$$E\left[X_{m,n}^{\max}\right] = E_{P_m}\left[\frac{X_{m,n}^{\max}}{X_m}\right],$$

where E_{P_m} denotes expectation with respect to P_m. For $m \le j \le n$ define $Y_j = X_j/X_m$. It is straightforward to check that $(Y_j, \mathcal{F}_j), j = m, m+1, \ldots, n$ is a non-negative martingale with respect to P_m. Therefore

$$\gamma\left(E\left[X_{m,n}^{\max}\right]\right) = \gamma\left(E_{P_m}\left[\max_{m \le j \le n} Y_j\right]\right) \le E_{P_m}[Y_n \ln Y_n]$$
$$= E_{P_m}\left[\frac{X_n}{X_m} \ln \frac{X_n}{X_m}\right] = E\left[X_n \ln \frac{X_n}{X_m}\right] = D(P_n \parallel P_m). \quad \Box$$

Again a similar inequality is satisfied for the minimum of a martingale, i.e.

$$\gamma\left(E\left[X_{m,n}^{\min}\right]\right) \le D(P_n \parallel P_m). \tag{15}$$

Inspired by Barron (1991) convergence of log bounded martingales can be proved as follows. Let $X_1, X_2, \ldots$ be a non-negative martingale. Without loss of generality we will assume that $E[X_n] = 1$. Then

$$E[X_n \ln X_n] - E[X_m \ln X_m] = D(P_n \parallel P_m). \tag{16}$$

Inequality (16) implies that $E[X_n \ln X_n]$ because information divergence is non-negative. Assume that $E[X_n \ln X_n]$ is bounded. Then $D(P_n \parallel P_m)$ converges to 0 for m, n tending to infinity. Now Theorem 5 implies that

$$E\left[X_{m,n}^{\max} - X_{m,n}^{\min}\right] \to 0 \quad \text{for } m, n \to \infty. \tag{17}$$

Thus X_n is a Cauchy sequence in $L^1(\Omega, Q)$ and converges. Using Markov's inequality together with (17) gives

$$Q\left(X_{m,n}^{\max} - X_{m,n}^{\min} \ge \varepsilon\right) \to 0 \quad \text{for } m, n \to \infty$$

so that X_n is a Cauchy sequence with probability one. Therefore the martingale also converges pointwise almost surely. Thus if $E[X_n \ln (X_n)]$ is bounded we get both mean and almost sure pointwise convergence.

By a similar argument, both mean convergence and almost sure pointwise convergence of

$$\frac{\mathrm{d}P_n}{\mathrm{d}Q} = \frac{1}{k} \sum_{j=0}^{k-1} f \circ T^j \tag{18}$$

hold for an ergodic transformation T when $E[|f| \log |f|] < \infty$.

In Cover and Thomas (1991) both convergence of martingales and stationary sequences are used in the proof of the Shannon–McMillan–Breiman theorem. It is interesting that exactly the finiteness of $E[X \ln X]$ (finite entropy) is needed in this theorem.
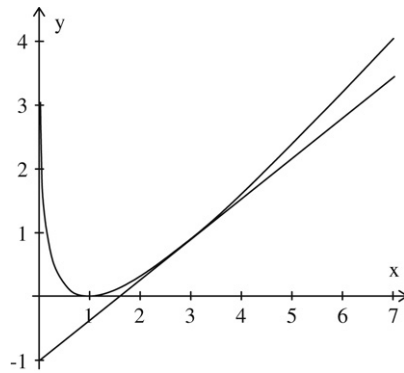
Fig. 1. The curve is the graph of the function γ. Doob's bound corresponds to the tangent.

4. Discussion

Theorem 3 can be seen as a strengthening of a classical maximal inequality by Doob, which states that

$$E[X^{\max}] \leq \frac{e}{e-1}\left(1 + E[X_n \ln(X_n)]\right). \tag{19}$$

As illustrated in Fig. 1 Doob's inequality corresponds to a tangent to the function $x \curvearrowright \gamma(x)$ at $x = e$. Thus, the new inequality is superior to Doob's inequality in a neighborhood of 1, and it is the behavior in this region that implies convergence of the martingale. Only in a neighborhood of $E[X^{\max}] = e$ Doob's inequality and the new inequalities gives approximately the same bound on $E[X^{\max}]$.

In this paper upper bounds on $E[X^{\max}]$ and lower bounds on $E[X^{\min}]$ are given in terms of $E[X_n \ln(X_n)]$, and each of the bounds is shown to be tight. In Theorem 4 the tightness of upper and lower bounds are obtained for different values of the parameter α. Therefore a tighter bound on $E[X^{\max} - X^{\min}]$ is possible in terms of $E[X_n \ln(X_n)]$. Such tighter bounds would be highly interesting and are obvious subjects for further investigation.

In Barron (1991) Pinsker's inequality was used to see that convergence in information implies convergence in total variation. If $P \ll Q$ then the sequence $(1, \frac{dP}{dQ})$ is a martingale. We have

$$E\left[\max\left\{1, \frac{dP}{dQ}\right\}\right] = 1 + \frac{1}{2}\|P - Q\|, \tag{20}$$

where the norm denotes variational distance. Then inequality (7) states that

$$\frac{1}{2}\|P - Q\| - \ln\left(1 + \frac{1}{2}\|P - Q\|\right) \leq D(P \| Q). \tag{21}$$

This inequality is well known and dates back to Volkonskij and Rozanov (1959) and was later refined to Pinsker's inequality, see Fedotov et al. (2003) for more details about the history of this problem. If the minimum is used rather than the maximum one gets an inequality that in some cases is stronger than the well-known Pinsker inequality.

Acknowledgements

References

Barron, A., Information theory and martingales. In: Proceedings International Symposium on Information Theory, Budapest; 1991.
Cover, T., Thomas, J.A., 1991. Elements of Information Theory. Wiley.
Csiszár, I., 1963. Eine informationstheoretische Ungleichung und ihre anwendung auf den Beweis der ergodizität von Markoffschen Ketten. Publ. Math. Inst. Hungar. Acad. 8, 95–108.
Csiszár, I., Shields, P., 2004. Information Theory and Statistics: A Tutorial. In: Foundations and Trends in Communications and Information Theory, Now Publishers Inc.
Fedotov, A., Harremoës, P., Topsøe, F., 2003. Refinements of Pinsker's Inequality. IEEE Trans. Inform. Theory 49 (6), 1491–1498.
Shiryaev, A.N., 1996. Probability. Springer, New York.
Volkonskij, V.A., Rozanov, J.A., 1959. Some limit theorems for random functions — I. Theory Probab. Appl. 4, 178–197.