Some New Maximal Inequalities

Peter Harremoës

Centrum voor Wiskunde en Informatica, Amsterdam, The Nederlands

Abstract

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

Keywords: Martingale, maximal inequality, stationary sequence.

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 us-

[1] Kruislaan 413, PO Box 94079, 1098 SJ Amsterdam *E-mail:* P.Harremoes@cwi.nl
Supported by the Danish Natural Science Research Council.

ing information theoretic techniques, see (Csiszár 1963), (Barron 1991) and (Csiszár & Shields 2004). Large deviation bounds are closely related to information theory and can be used to prove the strong law of large numbers (Csiszár & 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 all mean values will be calculated 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) *Assume that $X_1, X_2, ..., X_n$ is a sequence of non-negative random variables that form a 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*

$$\lambda \cdot Q\left(X^{\max} \geq \lambda\right) \leq E\left(X_n \cdot 1_{X^{\max} \geq \lambda}\right) \text{ and } \lambda \cdot Q\left(X^{\min} \leq \lambda\right) \geq E\left(X_n \cdot 1_{X^{\min} \leq \lambda}\right). \tag{1}$$

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*

$$f_n^{\min} = \min_{1 \le k \le n} \frac{1}{k} \sum_{j=0}^{k-1} f \circ T^j \text{ and } f_n^{\max} = \max_{1 \le k \le n} \frac{1}{k} \sum_{j=0}^{k-1} f \circ T^j. \tag{2}$$

Then

$$\lambda \cdot Q\left(f_n^{\max} \ge \lambda\right) \le E\left(f \cdot 1_{f_n^{\max} \ge \lambda}\right) \text{ and } \lambda \cdot Q\left(f_n^{\min} \le \lambda\right) \ge E\left(f \cdot 1_{f_n^{\min} \le \lambda}\right). \tag{3}$$

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* $X_1, X_2, ..., X_n$ *be a non-negative martingale. Let* $X^{\max} = \max(X_j)$ *and* $X^{\min} = \min(X_j)$*. If* $X_1 = 1$ *then*

$$\gamma\left(E\left(X^{\max}\right)\right) \le E\left(X_n \ln\left(X_n\right)\right) \tag{4}$$

and

$$\gamma\left(E\left(X^{\min}\right)\right) \le E\left(X_n \ln\left(X_n\right)\right). \tag{5}$$

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

$$
\begin{aligned}
E\left(X^{\max}\right) - 1 &= \int_0^\infty Q\left(X^{\max} \geq t\right)\ dt - 1 = \int_1^\infty Q\left(X^{\max} \geq t\right)\ dt \qquad (6)\\
&\leq \int_1^\infty \frac{1}{t} E\left(X_n \cdot 1_{X^{\max} \geq t}\right)\ dt = E\left(\int_1^\infty \frac{X_n \cdot 1_{X^{\max} \geq t}}{t}\ dt\right)\\
&= E\left(X_n \cdot \int_1^{X^{\max}} \frac{1}{t}\ dt\right) = E\left(X_n \cdot \ln\left(X^{\max}\right)\ dt\right).
\end{aligned}
$$

Using that γ is nonnegative we get

$$
\begin{aligned}
E\left(X^{\max}\right) - 1 &\leq E\left(X_n\left(\ln\left(X^{\max}\right) + \gamma\left(\frac{X^{\max}}{X_n \cdot E\left(X^{\max}\right)}\right)\right)\right) \qquad (7)\\
&= E\left(X_n \ln\left(X_n\right)\right) + \ln E\left(X^{\max}\right).
\end{aligned}
$$

Inequality (4) is obtained by reorganizing the terms.

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

$$
\begin{aligned}
E\left(X^{\min}\right) &= \int_0^1 Q\left(X^{\min} \geq t\right)\ dt = 1 - \int_0^1 Q\left(X^{\min} < t\right)\ dt \qquad (8)\\
&\leq 1 - \int_0^1 \frac{1}{t} E\left(X_n \cdot 1_{X^{\min} < t}\right)\ dt = 1 - E\left(\int_0^1 \frac{X_n \cdot 1_{X^{\min} < t}}{t}\ dt\right)\\
&= 1 - E\left(X_n \cdot \int_{X^{\min}}^1 \frac{1}{t}\ dt\right) = 1 + E\left(X_n \cdot \ln\left(X^{\min}\right)\right).
\end{aligned}
$$

Inequality (5) can now be proved in the same way as (4).

The theorem easily generalizes to martingales with continuous time. A similar inequality holds for an ergodic transformation T and a non-negative function f with $\int f\ dQ = 1$. Next we shall see that the inequalities in Theorem 3) are tight. Thus for each point on the curve $y = \gamma(x)$ we have to find a martingale such that (4) or (5) holds with equality. To construct such a martingale the time has to be continuous. Let Q be the uniform distribution on $[0;1]$. Let $\mathbb{F}_t$ be the σ-algebra generated by the Borel measurable subsets on $[0;t]$ and the set $[t;1]$. Let f denote the function (random variable) on $[0;1]$ given by $f(x) = (\beta+1)(1-x)^{\beta}$ where $\beta \in \left]-1;0\right]$. Remark that f is increasing. The conditional expectation of f given $\mathbb{F}_t$ is

$$\mathbb{E}(f \mid \mathbb{F}_t)(x) = \begin{cases} f(x) & \text{for } x < t, \\ \frac{\int_t^1 f(y)\ dy}{1-t} & \text{for } x \geq t. \end{cases} \tag{9}$$

Let $f^{\max} = \sup_{t\in[0;1]} \mathbb{E}(f \mid \mathbb{F}_t)(x)$. The supremum is attained for $t = x$ as illustrated on Figure 1. Thus

$$f^{\max}(x) = \frac{\int_x^1 f(y)\ dy}{1-x} = \frac{\int_x^1 (\beta+1)(1-y)^{\beta}\ dy}{1-x} = \frac{f(x)}{\beta+1}. \tag{10}$$

This implies that $f^{\max}/\left(f \cdot \int f^{\max}\ dQ\right) = 1$ and that Lemma 1 and Inequality (7) holds with equality. Therefore also Inequality (4) holds with equality and Inequality (4) is tight for $E(f^{\max}) = (\beta+1)^{-1}$. This holds for all $\beta \in \left]-1;0\right]$ and thus for all values of $E(f^{\max}) \in \left[1;\infty\right[$.

If β is positive in the above example then f is decreasing and $f^{\max}$ shall be replaced by $f^{\min}$. All the calculations are the same, and therefore also

Inequality (5) is tight.

3 Convergence of martingales and ergodic sequences

In order to prove 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 $D(P\|Q) = \int \ln \frac{dP}{dQ}\, dP$ if P is absolutely continuous with respect to Q and by $D(P\|Q) = \infty$ otherwise.

Theorem 4 *Let $X_1, X_2, ...$ be a non-negative martingale and assume that $E(X_j) = 1$. Let P_j be the probability measure given by $\frac{dP_j}{dQ} = X_j$. For $m \leq n$ put $X_{m,n}^{\max} = \sup_{j=m,\ldots,n} X_j$ and Let $\gamma(x) = x - 1 - \ln(x)$. Then*

$$\gamma\left(E\left(X_{m,n}^{\max}\right)\right) \leq D(P_n\|P_m). \tag{11}$$

PROOF. For each value x of X_m we have

$$\gamma\left(E\left(\frac{X^*}{X_m} \mid X_m = x\right)\right) \leq E\left(\frac{X_n}{X_m}\ln\left(\frac{X_n}{X_m}\right) \mid X_m = x\right). \tag{12}$$

Using convexity of γ leads to

$$\begin{aligned} E\left(X_m\gamma\left(E\left(\frac{X^*}{X_m} \mid X_m\right)\right)\right) &\leq E\left(X_m E\left(\frac{X_n}{X_m}\ln\left(\frac{X_n}{X_m}\right) \mid X_m\right)\right) \qquad (13) \\ &= E\left(X_n \ln\frac{X_n}{X_m}\right) = D(P_n\|P_m). \end{aligned}$$

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

$$\gamma\left(E\left(X_{m,n}^{\min}\right)\right) \leq D\left(P_n \| P_m\right). \tag{14}$$

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

$$E\left(X_n \ln X_n\right) - E\left(X_m \ln X_m\right) = D\left(P_n \| P_m\right). \tag{15}$$

We see that $E\left(X_n \ln X_n\right)$ is increasing. Assume that $E\left(X_n \ln X_n\right)$ is bounded. Then $D\left(P_n \| P_m\right)$ converges to 0 for m, n tending to infinity. Now Theorem 4 implies that $E\left(X_{m,n}^{\max} - X_{m,n}^{\min}\right) \to 0$ for $m, n \to \infty$. Thus X_n is a Cauchy sequence in $L^1\left(\Omega, Q\right)$ and converges. Further,

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

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

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

$$\frac{dP_n}{dQ} = \frac{1}{k} \sum_{j=0}^{k-1} f \circ T^j \tag{16}$$

for an ergodic transformation T when $E(|f| \log(|f|)) < \infty$.

In (Cover & 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.

4 Discussion

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

$$E(X^{\max}) \leq \frac{e}{e-1}\left(1 + E(X_n \ln(X_n))\right). \tag{17}$$

As illustrated on Figure 2 Doob's inequality corresponds to a tangent to the function γ. Thus the new inequality is superior to Doob's inequality in a neighborhood of 1, and it the behavior in this region which implies convergence of the martingale. Only in a neighborhood of $E(X^{\max}) = e$ Doob's inequality is optimal.

In this paper upper bounds for $E(X^{\max})$ and lower bounds on $E\left(X^{\min}\right)$ are given in terms of $E(X_n \ln(X_n))$, and each of the bounds is shown to be tight. In the example the tightness of upper and lower bounds are obtained for different values of the parameter β. Therefore a tighter bound on $E\left(X^{\max} - X^{\min}\right)$

is possible in terms of $E\left(X_n \ln\left(X_n\right)\right)$. Such tighter bounds would be highly interesting and are an obvious subject 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 $\left(1, \frac{dP}{dQ}\right)$ is a martingale. We have

$$E\left(\max\left\{1, \frac{dP}{dQ}\right\}\right) = 1 + \frac{1}{2}\left\|P - Q\right\|, \tag{18}$$

where the norm is the total variation norm. Then Inequality (4) states that

$$\frac{1}{2}\left\|P - Q\right\| - \ln\left(1 + \frac{1}{2}\left\|P - Q\right\|\right) \leq D\left(P\|Q\right). \tag{19}$$

This inequality is well-known and dates back to (Volkonskij & Rozanov 1959) and was later refined to Pinsker's inequality, see (Fedotov, Harremoës & Topsøe 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.

References

Barron, A. (1991). Information theory and martingales, *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*, 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 Prob. Appl.* **4**: 178 – 197.

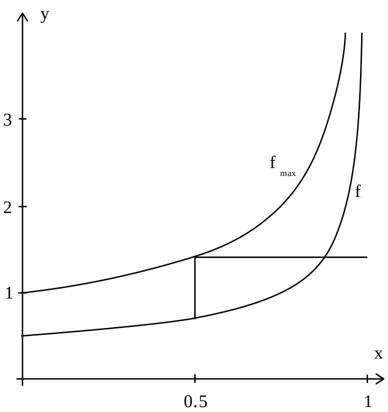

Fig. 1. For $\beta = -1/2$ the function f (lower curve) and $f_{\max}$ (upper curve) are illustrated. The conditional expectation $\mathbb{E}\left(f \mid \mathbb{F}_{1/2}\right)(x)$ is constant for $x > 1/2$ and equals f for $x \leq 1/2$.

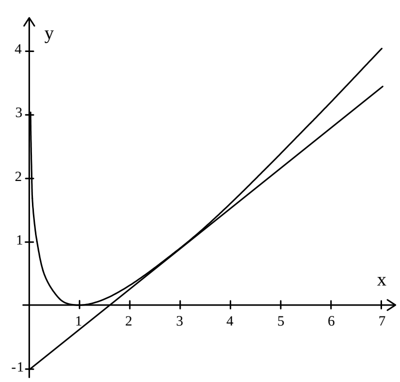

Fig. 2. The curve is graf of the function γ. Doob's bound corresponds to the tangent.