Indefinitely Oscillating Martingales

Jan Leike and Marcus Hutter

The Australian National University
{jan.leike,marcus.hutter}@anu.edu.au

Abstract. We construct a class of nonnegative martingale processes that oscillate indefinitely with high probability. For these processes, we state a uniform rate of the number of oscillations for a given magnitude and show that this rate is asymptotically close to the theoretical upper bound. These bounds on probability and expectation of the number of upcrossings are compared to classical bounds from the martingale literature. We discuss two applications. First, our results imply that the limit of the minimum description length operator may not exist. Second, we give bounds on how often one can change one's belief in a given hypothesis when observing a stream of data.[1]

Keywords: Martingales, infinite oscillations, bounds, convergence rates, minimum description length, mind changes.

1 Introduction

Martingale processes model fair gambles where knowledge of the past or choice of betting strategy have no impact on future winnings. But their application is not restricted to gambles and stock markets. Here we exploit the connection between nonnegative martingales and probabilistic data streams, i.e., probability measures on infinite strings. For two probability measures P and Q on infinite strings, the quotient Q/P is a nonnegative P-martingale. Conversely, every nonnegative P-martingale is a multiple of Q/P P-almost everywhere for some probability measure Q.

One of the famous results of martingale theory is Doob's Upcrossing Inequality [Doo53]. The inequality states that in expectation, every nonnegative martingale has only finitely many oscillations (called *upcrossings* in the martingale literature). Moreover, the bound on the expected number of oscillations is inversely proportional to their magnitude. Closely related is Dubins' Inequality [Dur10] which asserts that the probability of having many oscillations decreases exponentially with their number. These bounds are given with respect to oscillations of fixed magnitude.

In Section 4 we construct a class of nonnegative martingale processes that have infinitely many oscillations of (by Doob necessarily) decreasing magnitude.

[1] In Theorem 4, Q needs to be absolutely continuous with respect to P on cylinder sets. In Theorem 6, Corollary 7, Corollary 8, and Corollary 13, P needs to have perpetual entropy. See technical report [LH14].

P. Auer et al. (Eds.): ALT 2014, LNAI 8776, pp. 321–335, 2014.

These martingales satisfy uniform lower bounds on the probability and the expectation of the number of upcrossings. We prove corresponding upper bounds in Section 5 showing that these lower bounds are asymptotically tight. Moreover, the construction of the martingales is agnostic regarding the underlying probability measure, assuming only mild restrictions on it. We compare these results to the statements of Dubins' Inequality and Doob's Upcrossing Inequality and demonstrate that our process makes those inequalities (in Doob's case asymptotically) tight. If we drop the uniformity requirement, asymptotics arbitrarily close to Doob and Dubins' bounds are achievable. We discuss two direct applications of these bounds.

The Minimum Description Length (MDL) principle [Ris78] and the closely related Minimal Message Length (MML) principle [WB68] recommend to select among a class of models the one that has the shortest code length for the data plus code length for the model. There are many variations, so the following statements are generic: for a variety of problem classes MDL's predictions have been shown to converge asymptotically (predictive convergence). For continuous independently identically distributed data the MDL estimator usually converges to the true distribution [Grü07, Wal05] (inductive consistency). For arbitrary (non-i.i.d.) countable classes, the MDL estimator's predictions converge to those of the true distribution for single-step predictions [PH05] and ∞-step predictions [Hut09]. Inductive consistency implies predictive convergence, but not the other way around. In Section 6 we show that indeed, the MDL estimator for countable classes is *inductively inconsistent*. This can be a major obstacle for using MDL for prediction, since the model used for prediction has to be changed over and over again, incurring the corresponding computational cost.

Another application of martingales is in the theory of mind changes [LS05]. How likely is it that your belief in some hypothesis changes by at least $\alpha > 0$ several times while observing some evidence? Davis recently showed [Dav13] using elementary mathematics that this probability decreases exponentially. In Section 7 we rephrase this problem in our setting: the stochastic process

$$P(\,\text{hypothesis} \mid \text{evidence up to time } t\,)$$

is a martingale bounded between 0 and 1. The upper bound on the probability of many changes can thus be derived from Dubins' Inequality. This yields a simpler alternative proof for Davis' result. However, because we consider nonnegative but unbounded martingales, we get a weaker bound than Davis.

Omitted proofs can be found in the extended technical report [LH14].

2 Strings, Measures, and Martingales

We presuppose basic measure and probability theory [Dur10, Chp.1]. Let Σ be a finite set, called *alphabet*. We assume Σ contains at least two distinct elements. For every $u \in \Sigma^*$, the *cylinder set*

$$\Gamma_u := \{uv \mid v \in \Sigma^\omega\}$$

is the set of all infinite strings of which u is a prefix. Furthermore, fix the σ-algebras

$$\mathcal{F}_t := \sigma\left(\{\Gamma_u \mid u \in \Sigma^t\}\right) \qquad \text{and} \qquad \mathcal{F}_\omega := \sigma\Big(\bigcup_{t=1}^{\infty} \mathcal{F}_t\Big).$$

$(\mathcal{F}_t)_{t\in\mathbb{N}}$ is a *filtration*: since $\Gamma_u = \bigcup_{a\in\Sigma} \Gamma_{ua}$, it follows that $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$ for every $t \in \mathbb{N}$, and all $\mathcal{F}_t \subseteq \mathcal{F}_\omega$ by the definition of $\mathcal{F}_\omega$. An *event* is a measurable set $E \subseteq \Sigma^\omega$. The event $E^c := \Sigma^\omega \setminus E$ denotes the complement of E.

Definition 1 (Stochastic Process). $(X_t)_{t\in\mathbb{N}}$ *is called* ($\mathbb{R}$-valued) stochastic process *iff each* X_t *is an* $\mathbb{R}$*-valued random variable.*

Definition 2 (Martingale). *Let* P *be a probability measure over* $(\Sigma^\omega, \mathcal{F}_\omega)$*. An* $\mathbb{R}$*-valued stochastic process* $(X_t)_{t\in\mathbb{N}}$ *is called a* P-supermartingale (P-submartingale) *iff*

(a) each X_t *is* $\mathcal{F}_t$*-measurable, and*
(b) $\mathbb{E}[X_t \mid \mathcal{F}_s] \leq X_s$ $(\mathbb{E}[X_t \mid \mathcal{F}_s] \geq X_s)$ *almost surely for all* $s, t \in \mathbb{N}$ *with* $s < t$*.*

A process that is both P*-supermartingale and* P*-submartingale is called* P-martingale.

We call a supermartingale (submartingale) process $(X_t)_{t\in\mathbb{N}}$ *nonnegative* iff $X_t \geq 0$ for all $t \in \mathbb{N}$.

A *stopping time* is an $(\mathbb{N} \cup \{\omega\})$-valued random variable T such that $\{v \in \Sigma^\omega \mid T(v) = t\} \in \mathcal{F}_t$ for all $t \in \mathbb{N}$. Given a supermartingale $(X_t)_{t\in\mathbb{N}}$, the *stopped process* $(X_{\min\{t,T\}})_{t\in\mathbb{N}}$ is a supermartingale [Dur10, Thm. 5.2.6]. If $(X_t)_{t\in\mathbb{N}}$ is bounded, the limit of the stopped process, X_T, exists almost surely even if $T = \omega$ (Martingale Convergence Theorem [Dur10, Thm. 5.2.8]). We use the following variant on Doob's Optional Stopping Theorem for supermartingales.

Theorem 3 (Optional Stopping Theorem [Dur10, Thm. 5.7.6]). *Let* $(X_t)_{t\in\mathbb{N}}$ *be a nonnegative supermartingale and let* T *be a stopping time. The random variable* X_T *is almost surely well defined and* $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$*.*

We exploit the following two theorems that state the connection between probability measures on infinite strings and martingales. For any two probability measures P and Q on $(\Sigma^\omega, \mathcal{F}_\omega)$, the quotient Q/P is a nonnegative P-martingale. Conversely, for every nonnegative P-martingale there is a probability measure Q on $(\Sigma^\omega, \mathcal{F}_\omega)$ such that the martingale is P-almost surely a multiple of Q/P.

Theorem 4 (Measures → Martingales [Doo53, II§7 Ex. 3]). *Let* Q *and* P *be two probability measures on* $(\Sigma^\omega, \mathcal{F}_\omega)$*. The stochastic process* $(X_t)_{t\in\mathbb{N}}$*,* $X_t(v) := Q(\Gamma_{v_{1:t}})/P(\Gamma_{v_{1:t}})$ *is a nonnegative* P*-martingale with* $\mathbb{E}[X_t] = 1$*.*

Theorem 5 (Martingales → Measures [LH14]). *Let* P *be a probability measure on* $(\Sigma^\omega, \mathcal{F}_\omega)$ *and let* $(X_t)_{t\in\mathbb{N}}$ *be a nonnegative* P*-martingale with* $\mathbb{E}[X_t] = 1$*. There is a probability measure* Q *on* $(\Sigma^\omega, \mathcal{F}_\omega)$ *such that for all* $v \in \Sigma^\omega$ *and all* $t \in \mathbb{N}$ *with* $P(\Gamma_{v_{1:t}}) > 0$*,* $X_t(v) = Q(\Gamma_{v_{1:t}})/P(\Gamma_{v_{1:t}})$*.*

3 Martingale Upcrossings

Fix $c \in \mathbb{R}$, and let $(X_t)_{t\in\mathbb{N}}$ be a martingale over the probability space $(\Sigma^\omega, \mathcal{F}_\omega, P)$. Let $t_1 < t_2$. We say the process $(X_t)_{t\in\mathbb{N}}$ does an ε-*upcrossing* between t_1 and t_2 iff $X_{t_1} \leq c - \varepsilon$ and $X_{t_2} \geq c + \varepsilon$. Similarly, we say $(X_t)_{t\in\mathbb{N}}$ does an ε-*downcrossing* between t_1 and t_2 iff $X_{t_1} \geq c + \varepsilon$ and $X_{t_2} \leq c - \varepsilon$. Except for the first upcrossing, consecutive upcrossings always involve intermediate downcrossings. Formally, we define the stopping times

$$\begin{aligned} T_0(v) &:= 0, \\ T_{2k+1}(v) &:= \inf\{t > T_{2k}(v) \mid X_t(v) \leq c - \varepsilon\}, \text{ and} \\ T_{2k+2}(v) &:= \inf\{t > T_{2k+1}(v) \mid X_t(v) \geq c + \varepsilon\}. \end{aligned}$$

The $T_{2k}(v)$ denote the indices of upcrossings. We count the number of upcrossings by the random variable $U_t^X(c-\varepsilon, c+\varepsilon)$, where

$$U_t^X(c-\varepsilon, c+\varepsilon)(v) := \sup\{k \geq 0 \mid T_{2k}(v) \leq t\}$$

and $U^X(c-\varepsilon, c+\varepsilon) := \sup_{t\in\mathbb{N}} U_t^X(c-\varepsilon, c+\varepsilon)$ denotes the total number of upcrossings. We omit the superscript X if the martingale $(X_t)_{t\in\mathbb{N}}$ is clear from context.

The following notation is used in the proofs. Given a monotone decreasing function $f : \mathbb{N} \to [0, 1)$ and $m, k \in \mathbb{N}$, we define the events $E_{m,k}^{X,f}$ that denote that there are at least k-many $f(m)$-upcrossings:

$$E_{m,k}^{X,f} := \left\{v \in \Sigma^\omega \mid U^X(1 - f(m), 1 + f(m))(v) \geq k\right\}.$$

For all $m, k \in \mathbb{N}$ we have $E_{m,k}^{X,f} \supseteq E_{m,k+1}^{X,f}$ and $E_{m,k}^{X,f} \subseteq E_{m+1,k}^{X,f}$. Again, we omit X and f in the superscript if they are clear from context.

4 Indefinitely Oscillating Martingales

In this section we construct a class of martingales that has a high probability of doing an infinite number of upcrossings. The magnitude of the upcrossings decreases at a rate of a given summable function f (a function f is called *summable* iff it has finite L_1-norm, i.e., $\sum_{i=1}^{\infty} f(i) < \infty$), and the value of the martingale X_t oscillates back and forth between $1 - f(M_t)$ and $1 + f(M_t)$, where M_t denotes the number of upcrossings so far. The process has a monotone decreasing chance of escaping the oscillation.

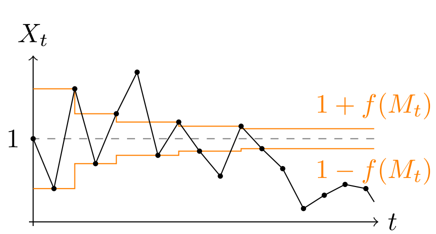

Fig. 1. An example evaluation of the martingale defined in the proof of Theorem 6

Theorem 6 (An Indefinitely Oscillating Martingale). *Let* $0 < \delta < \frac{2}{3}$ *and let* $f : \mathbb{N} \to [0, 1)$ *be any monotone decreasing function such that* $\sum_{i=1}^{\infty} f(i) \leq \frac{\delta}{2}$*. For every probability measure* P *with* $P(\Gamma_u) > 0$ *for all* $u \in \Sigma^*$ *there is a nonnegative martingale* $(X_t)_{t\in\mathbb{N}}$ *with* $\mathbb{E}[X_t] = 1$ *and*

$$P[\forall m.\, U(1 - f(m), 1 + f(m)) \geq m] \geq 1 - \delta.$$

Proof. We assume $\Sigma = \{0, 1\}$ by grouping symbols into two groups. Since $P(\Gamma_{u0} \mid \Gamma_u) + P(\Gamma_{u1} \mid \Gamma_u) = 1$, we can define a function $a : \Sigma^* \to \Sigma$ that assigns to every string $u \in \Sigma^*$ a symbol $a_u := a(u)$ such that $p_u := P(\Gamma_{ua_u} \mid \Gamma_u) \leq \frac{1}{2}$. By assumption, we have $p_u > 0$.

We define the following stochastic process $(X_t)_{t\in\mathbb{N}}$. Let $v \in \Sigma^\omega$ and $t \in \mathbb{N}$ be given and define $u := v_{1:t}$. Let

$$M_t(v) := 1 + \arg\max_{m\in\mathbb{N}} \left\{\forall k \leq m.\, U_t^X(1 - f(k), 1 + f(k)) \geq k\right\},$$

i.e., M_t is 1 plus the number of upcrossings completed up to time t. Define

$$\gamma_t(v) := \tfrac{p_u}{1-p_u}(1 + f(M_t(v)) - X_t(v)).$$

For $t = 0$, we set $X_0(v) := 1$, otherwise we distinguish the following three cases.

(i) For $X_t(v) \geq 1$:

$$X_{t+1}(v) := \begin{cases} 1 - f(M_t(v)) & \text{if } v_{t+1} \neq a_u, \\ X_t(v) + \frac{1-p_u}{p_u}(X_t(v) - (1 - f(M_t(v)))) & \text{if } v_{t+1} = a_u. \end{cases}$$

(ii) For $1 > X_t(v) \geq \gamma_t(v)$:

$$X_{t+1}(v) := \begin{cases} X_t(v) - \gamma_t(v) & \text{if } v_{t+1} \neq a_u, \\ 1 + f(M_t(v)) & \text{if } v_{t+1} = a_u. \end{cases}$$

(iii) For $X_t(v) < \gamma_t(v)$ and $X_t(v) < 1$:
let $d_t(v) := \max\{0, \min\{\frac{p_u}{1-p_u} X_t(v), \frac{1-p_u}{p_u}\gamma_t(v) - 2f(M_t(v))\}\}$;

$$X_{t+1}(v) := \begin{cases} X_t(v) + d_t(v) & \text{if } v_{t+1} \neq a_u, \\ X_t(v) - \frac{1-p_u}{p_u} d_t(v) & \text{if } v_{t+1} = a_u. \end{cases}$$

We give an intuition for the behavior of the process $(X_t)_{t\in\mathbb{N}}$. For all m, the following repeats. First X_t increases while reading a_u's until it reads one symbol that is not a_u and then jumps down to $1 - f(m)$. Subsequently, X_t decreases while not reading a_u's until it falls below γ_t or reads an a_u and then jumps up to $1 + f(m)$. If it falls below 1 and γ_t, then at every step, it can either jump up to $1 - f(m)$ or jump down to 0, whichever one is closest (the distance to the closest of the two is given by d_t). See Figure 1 for a visualization.

For notational convenience, in the following we omit writing the argument v to the random variables X_t, γ_t, M_t, and d_t.

Claim 1: $(X_t)_{t\in\mathbb{N}}$ is a martingale. Each X_{t+1} is $\mathcal{F}_{t+1}$-measurable, since it uses only the first $t+1$ symbols of v. Writing out cases (i), (ii), and (iii), we get

$$\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] \overset{(i)}{=} (1-f(M_t))(1-p_u) + \big(X_t + \tfrac{1-p_u}{p_u}(X_t - (1-f(M_t)))\big)p_u = X_t,$$

$$\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] \overset{(ii)}{=} \big(X_t - \tfrac{p_u}{1-p_u}((1+f(M_t)) - X_t)\big)(1-p_u) + (1+f(M_t))p_u = X_t,$$

$$\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] \overset{(iii)}{=} (X_t + d_t)(1-p_u) + (X_t - \tfrac{1-p_u}{p_u} d_t)p_u = X_t.$$

Claim 2: $X_t \geq 0$ and $\mathbb{E}[X_t] = 1$. The latter follows from $X_0 = 1$. Regarding the former, we use $0 \leq f(M_t) < 1$ to conclude

- (i$\neq$) $1 - f(M_t) \geq 0$,
- (i$=$) $\frac{1-p_u}{p_u}(X_t - (1-f(M_t))) \geq 0$ for $X_t \geq 1$,
- (ii$\neq$) $X_t - \gamma_t \geq 0$ for $X_t \geq \gamma_t$,
- (ii$=$) $1 + f(M_t) \geq 0$,
- (iii$\neq$) $X_t + d_t \geq 0$ since $d_t \geq 0$, and
- (iii$=$) $X_t - \frac{1-p_u}{p_u} d_t \geq 0$ since $d_t \leq \frac{p_u}{1-p_u} X_t$.

Claim 3: $X_t \leq 1 - f(M_t)$ or $X_t \geq 1 + f(M_t)$ for all $t \geq T_1$. We use induction on t. The induction start holds with $X_{T_1} \leq 1 - f(M_t)$. The induction step is clear for (i) $X_t \geq 1$ and (ii) $1 > X_t \geq \gamma_t$ since $\gamma_t \geq 0$. In case (iii) we have either $d_t = 0$ or $d_t \leq \frac{1-p_u}{p_u}\gamma_t - 2f(M_t)$ and since $X_t < \gamma_t$,

$$X_{t+1} \leq X_t + d_t \leq X_t + (1 + f(M_t) - X_t) - 2f(M_t) = 1 - f(M_t).$$

Claim 4: If $X_t \geq 1 - f(M_t)$ then $X_t > \gamma_t$. In this case

$$\gamma_t = \tfrac{p_u}{1-p_u}(1 + f(M_t) - X_t) \leq 2\tfrac{p_u}{1-p_u} f(M_t),$$

and thus with $p_u \leq \frac{1}{2}$ and $f(M_t) \leq \sum_{k=1}^{\infty} f(k) \leq \frac{\delta}{2} < \frac{1}{3}$,

$$X_t - \gamma_t \geq 1 - f(M_t) - 2\tfrac{p_u}{1-p_u} f(M_t) = 1 - \tfrac{1+p_u}{1-p_u} f(M_t) \geq 1 - 3f(M_t) > 0.$$

Claim 5: If $X_t > 0$ and ($f(M_t) > 0$ or $X_t < 1$) then $X_{t+1} \neq X_t$.

- (i) Assume $X_t \geq 1$. Then either $X_{t+1} = 1 - f(M_t) < 1$, or $\frac{1-p_u}{p_u}(X_t - (1 - f(M_t))) > 0$ since $X_t > 1 - f(M_t)$.
- (ii) Assume $1 > X_t \geq \gamma_t$. Then either $X_{t+1} = 1 + f(M_t) \geq 1 > X_t$, or $1 + f(M_t) - X_t \geq 1 - X_t > 0$, hence $\gamma_t > 0$ and thus $X_{t+1} = X_t - \gamma_t < X_t$.
- (iii) Assume $0 < X_t < \gamma_t$ and $X_t < 1$. From Claim 4 follows $X_t < 1 - f(M_t)$, thus $\frac{1-p_u}{p_u}\gamma_t - 2f(M_t) = 1 - f(M_t) - X_t > 0$. By assumption, $\frac{p_u}{1-p_u}X_t > 0$ and therefore $d_t > 0$. Hence $X_t + d_t > X_t$ and $X_t - \frac{1-p_u}{p_u} d_t < X_t$.

Claim 6: For all $m \in \mathbb{N}$, if $E_{m,m-1} \neq \emptyset$ then $P(E_{m,m} \mid E_{m,m-1}) \geq 1 - 2f(m)$. Let $v \in E_{m,m-1}$ and let $t_0 \in \mathbb{N}$ be a time step such that exactly $m-1$ upcrossings have been completed up to time t_0, i.e., $M_{t_0}(v) = m$. The subsequent downcrossing is completed eventually with probablity 1: we are in case (i) and in every step there is a chance of $1 - p_u \geq \frac{1}{2}$ of completing the downcrossing. Therefore we assume without loss of generality that the downcrossing has been completed, i.e., that t_0 is such that $X_{t_0}(v) = 1 - f(m)$. We will bound the probability $p := P(E_{m,m} \mid E_{m,m-1})$ that X_t rises above $1 + f(m)$ after t_0 to complete the m-th upcrossing.

Define the stopping time $T : \Sigma^\omega \to \mathbb{N} \cup \{\omega\}$,

$$T(v) := \inf\{t \geq t_0 \mid X_t(v) \geq 1 + f(m) \ \vee \ X_t(v) = 0\},$$

and define the stochastic process $Y_t = 1 + f(m) - X_{\min\{t_0+t,T\}}$. Because $(X_{\min\{t_0+t,T\}})_{t\in\mathbb{N}}$ is martingale, $(Y_t)_{t\in\mathbb{N}}$ is martingale. By definition, X_t always stops at $1 + f(m)$ before exceeding it, thus $X_T \leq 1 + f(m)$, and hence $(Y_t)_{t\in\mathbb{N}}$ is nonnegative. The Optional Stopping Theorem yields $\mathbb{E}[Y_{T-t_0} \mid \mathcal{F}_{t_0}] \leq \mathbb{E}[Y_0 \mid \mathcal{F}_{t_0}]$ and thus $\mathbb{E}[X_T \mid \mathcal{F}_{t_0}] \geq \mathbb{E}[X_{t_0} \mid \mathcal{F}_{t_0}] = 1 - f(m)$. By Claim 5, X_t does not converge unless it reaches either 0 or $1 + f(m)$, and thus

$$1 - f(m) \leq \mathbb{E}[X_T \mid \mathcal{F}_{t_0}] = (1 + f(m)) \cdot p + 0 \cdot (1 - p),$$

hence $P(E_{m,m} \mid E_{m,m-1}) = p \geq 1 - f(m)(1 + p) \geq 1 - 2f(m)$.

Claim 7: $E_{m+1,m} = E_{m,m}$ and $E_{m+1,m+1} \subseteq E_{m,m}$. By definition of M_t, the i-th upcrossings of the process $(X_t)_{t\in\mathbb{N}}$ is between $1 - f(i)$ and $1 + f(i)$. The function f is monotone decreasing, and by Claim 3 the process $(X_t)_{t\in\mathbb{N}}$ does not assume values between $1 - f(i)$ and $1 + f(i)$. Therefore the first m $f(m + 1)$-upcrossings are also $f(m)$-upcrossings, i.e., $E_{m+1,m} \subseteq E_{m,m}$. By definition of $E_{m,k}$ we have $E_{m+1,m} \supseteq E_{m,m}$ and $E_{m+1,m+1} \subseteq E_{m+1,m}$.

Claim 8: $P(E_{m,m}) \geq 1 - \sum_{i=1}^m 2f(i)$. For $P(E_{0,0}) = 1$ this holds trivially. Using Claim 6 and Claim 7 we conclude inductively

$$\begin{aligned} P(E_{m,m}) &= P(E_{m,m} \cap E_{m,m-1}) = P(E_{m,m} \mid E_{m,m-1})P(E_{m,m-1}) \\ &= P(E_{m,m} \mid E_{m,m-1})P(E_{m-1,m-1}) \\ &\geq (1 - 2f(m)) \left(1 - \sum_{i=1}^{m-1} 2f(i)\right) \geq 1 - \sum_{i=1}^{m} 2f(i). \end{aligned}$$

From Claim 7 follows $\bigcap_{i=1}^m E_{i,i} = E_{m,m}$ and therefore $P(\bigcap_{i=1}^\infty E_{i,i}) = \lim_{m\to\infty} P(E_{m,m}) \geq 1 - \sum_{i=1}^\infty 2f(i) \geq 1 - \delta$. □

Theorem 6 gives a *uniform* lower bound on the probability for many upcrossings: it states the probability of the event that *for all* $m \in \mathbb{N}$, $U(1 - f(m), 1 + f(m)) \geq m$ holds. This is a lot stronger than the nonuniform bound $P[U(1 - f(m), 1 + f(m)) \geq m] \geq 1 - \delta$ for all $m \in \mathbb{N}$: the quantifier is inside the probability statement.

As an immediate consequence of Theorem 6, we get the following uniform lower bound on the *expected* number of upcrossings.

Corollary 7 (Expected Upcrossings). *Under the same conditions as in Theorem 6, for all $m \in \mathbb{N}$,*

$$\mathbb{E}[U(1 - f(m), 1 + f(m))] \geq m(1 - \delta).$$

Proof. From Theorem 6 and Markov's inequality. □

By choosing the slowly decreasing but summable function f by setting $f^{-1}(\varepsilon) := 2\delta(\frac{1}{\varepsilon(\ln\varepsilon)^2} - \frac{e^2}{4})$, we get the following concrete results.

Corollary 8 (Concrete Lower Bound). *Let* $0 < \delta < 1$. *For every probability measure* P *with* $P(\Gamma_u) > 0$ *for all* $u \in \Sigma^*$, *there is a nonnegative martingale* $(X_t)_{t\in\mathbb{N}}$ *with* $\mathbb{E}[X_t] = 1$ *such that*

$$P\left[\forall\varepsilon > 0.\, U(1-\varepsilon, 1+\varepsilon) \in \Omega\left(\frac{\delta}{\varepsilon\left(\ln\frac{1}{\varepsilon}\right)^2}\right)\right] \geq 1-\delta \text{ and}$$
$$\mathbb{E}[U(1-\varepsilon, 1+\varepsilon)] \in \Omega\Big(\frac{1}{\varepsilon\left(\ln\frac{1}{\varepsilon}\right)^2}\Big).$$

Moreover, for all $\varepsilon < 0.015$ *we get* $\mathbb{E}[U(1-\varepsilon, 1+\varepsilon)] > \frac{\delta(1-\delta)}{\varepsilon\left(\ln\frac{1}{\varepsilon}\right)^2}$ *and*

$$P\left[\forall\varepsilon < 0.015.\, U(1-\varepsilon, 1+\varepsilon) > \frac{\delta}{\varepsilon\left(\ln\frac{1}{\varepsilon}\right)^2}\right] \geq 1-\delta.$$

The concrete bounds given in Theorem 8 are *not* the asymptotically optimal ones: there are summable functions that decrease even more slowly. For example, we could multiply f^{-1} with the factor $\sqrt{\ln(1/\varepsilon)}$ (which still is not optimal).

5 Martingale Upper Bounds

In this section we state upper bounds on the probability and expectations of many upcrossings (Dubins' Inequality and Doob's Upcrossing Inequality). We use the construction from the previous section to show that these bounds are tight. Moreover, with the following theorem we show that the uniform lower bound on the probability of many upcrossings guaranteed in Theorem 6 is asymptotically tight.

Every function f is either summable or not. If f is summable, then we can scale it with a constant factor such that its sum is smaller than $\frac{\delta}{2}$, and then apply the construction of Theorem 6. If f is not summable, the following theorem implies that there is no *uniform* lower bound on the probability of having at least m-many $f(m)$-upcrossings.

Theorem 9 (Upper Bound on Upcrossing Rate). *Let* $f : \mathbb{N} \to [0, 1)$ *be a monotone decreasing function such that* $\sum_{t=1}^{\infty} f(t) = \infty$. *For every probability measure* P *and for every nonnegative* P*-martingale* $(X_t)_{t\in\mathbb{N}}$ *with* $\mathbb{E}[X_t] = 1$,

$$P[\forall m.\, U(1-f(m), 1+f(m)) \geq m] = 0.$$

Proof. Define the events $D_m := \bigcup_{i=1}^m E_{i,i}^c = \{\forall i \leq m.\, U(1-f(i), 1+f(i)) \geq i\}$. Then $D_m \subseteq D_{m+1}$. Assume there is a constant $c > 0$ such that $c \leq P(D_m^c) = P(\bigcap_{i=1}^m E_{i,i})$ for all m. Let $m \in \mathbb{N}$, $v \in D_m^c$, and pick $t_0 \in \mathbb{N}$ such that the process $X_0(v), \ldots, X_{t_0}(v)$ has completed i-many $f(i)$-upcrossings for all $i \leq m$

and $X_{t_0}(v) \leq 1 - f(m+1)$. If $X_t(v) \geq 1 + f(m+1)$ for some $t \geq t_0$, the $(m+1)$-st upcrossing for $f(m+1)$ is completed and thus $v \in E_{m+1,m+1}$. Define the stopping time $T : \Sigma^\omega \to (\mathbb{N} \cup \{\omega\})$,

$$T(v) := \inf\{t \geq t_0 \mid X_t(v) \geq 1 + f(m+1)\}.$$

According to the Optional Stopping Theorem applied to the process $(X_t)_{t \geq t_0}$, the random variable X_T is almost surely well-defined and $\mathbb{E}[X_T \mid \mathcal{F}_{t_0}] \leq \mathbb{E}[X_{t_0} \mid \mathcal{F}_{t_0}] = X_{t_0}$. This yields $1 - f(m+1) \geq X_{t_0} \geq \mathbb{E}[X_T \mid \mathcal{F}_{t_0}]$ and by taking the expectation $\mathbb{E}[\,\cdot \mid X_{t_0} \leq 1 - f(m+1)]$ on both sides,

$$\begin{aligned} 1 - f(m+1) &\geq \mathbb{E}[X_T \mid X_{t_0} \leq 1 - f(m+1)] \\ &\geq (1 + f(m+1)) P[X_T \geq 1 + f(m+1) \mid X_{t_0} \leq 1 - f(m+1)] \end{aligned}$$

by Markov's inequality. Therefore

$$\begin{aligned} P(E_{m+1,m+1} \mid D_m^c) &= P[X_T \geq 1 + f(m+1) \mid X_{t_0} \leq 1 - f(m+1)] \\ &\quad \cdot P[X_{t_0} \leq 1 - f(m+1) \mid D_m^c] \\ &\leq P[X_T \geq 1 + f(m+1) \mid X_{t_0} \leq 1 - f(m+1)] \\ &\leq \tfrac{1 - f(m+1)}{1 + f(m+1)} \leq 1 - f(m+1). \end{aligned}$$

Together with $c \leq P(D_m^c)$ we get

$$\begin{aligned} P\left(D_{m+1} \setminus D_m\right) &= P\left(E_{m+1,m+1}^c \cap D_m^c\right) \\ &= P\left(E_{m+1,m+1}^c \mid D_m^c\right) P\left(D_m^c\right) \geq f(m+1)c. \end{aligned}$$

This is a contradiction because $\sum_{i=1}^\infty f(i) = \infty$:

$$1 \geq P(D_{m+1}) = P\left(\biguplus_{i=1}^m (D_{i+1} \setminus D_i)\right) = \sum_{i=1}^m P(D_{i+1} \setminus D_i) \geq \sum_{i=1}^m f(i+1)c \to \infty.$$

Therefore the assumption $P(D_m^c) \geq c$ for all m is false, and hence we get $P[\forall m.\, U(1 - f(m), 1 + f(m)) \geq m] = P(\bigcap_{i=1}^\infty E_{i,i}) = \lim_{m \to \infty} P(D_m^c) = 0$. □

By choosing the decreasing non-summable function f by setting $f^{-1}(\varepsilon) := \frac{-a}{\varepsilon(\ln \varepsilon)} - b$ for Theorem 9, we get that $U(1 - \varepsilon, 1 + \varepsilon) \notin \Omega(\frac{1}{\varepsilon \log(1/\varepsilon)})$ P-almost surely.

Corollary 10 (Concrete Upper Bound). *Let P be a probability measure and let $(X_t)_{t \in \mathbb{N}}$ be a nonnegative martingale with $\mathbb{E}[X_t] = 1$. Then for all $a, b > 0$,*

$$P\left[\forall \varepsilon > 0.\, U(1 - \varepsilon, 1 + \varepsilon) \geq \tfrac{a}{\varepsilon \log(1/\varepsilon)} - b\right] = 0.$$

Theorem 11 (Dubins' Inequality [Dur10, Ex.5.2.14]). *For every nonnegative P-martingale $(X_t)_{t \in \mathbb{N}}$ and for every $c > 0$ and every $\varepsilon > 0$,*

$$P[U(c - \varepsilon, c + \varepsilon) \geq k] \leq \left(\tfrac{c - \varepsilon}{c + \varepsilon}\right)^k \mathbb{E}\left[\min\left\{\tfrac{X_0}{c - \varepsilon}, 1\right\}\right].$$

Dubins' Inequality immediately yields the following bound on the probability of the number of upcrossings.

$$P[U(1-f(m), 1+f(m)) \geq k] \leq \left(\tfrac{1-f(m)}{1+f(m)}\right)^k.$$

The construction from Theorem 6 shows that this bound is asymptotically tight for $m = k \to \infty$ and $\delta \to 0$: define the monotone decreasing function $f : \mathbb{N} \to [0, 1)$,

$$f(t) := \begin{cases} \frac{\delta}{2k}, & \text{if } t \leq k, \text{ and} \\ 0, & \text{otherwise.} \end{cases}$$

Then the martingale from Theorem 6 yields the lower bound

$$P[U(1-\tfrac{\delta}{2k}, 1+\tfrac{\delta}{2k}) \geq k] \geq 1-\delta,$$

while Dubins' Inequality gives the upper bound

$$P[U(1-\tfrac{\delta}{2k}, 1+\tfrac{\delta}{2k}) \geq k] \leq \left(\frac{1-\frac{\delta}{2k}}{1+\frac{\delta}{2k}}\right)^k = \left(1-\frac{2\delta}{2k+\delta}\right)^k \xrightarrow{k\to\infty} \exp(-\delta).$$

As δ approaches 0, the value of $\exp(-\delta)$ approaches $1-\delta$ (but exceeds it since exp is convex). For $\delta = 0.2$ and $m = k = 3$, the difference between the two bounds is already lower than 0.021.

The following theorem places an upper bound on the rate of *expected* upcrossings.

Theorem 12 (Doob's Upcrossing Inequality [Xu12]). *Let $(X_t)_{t\in\mathbb{N}}$ be a submartingale. For every $c \in \mathbb{R}$ and $\varepsilon > 0$,*

$$\mathbb{E}[U_t(c-\varepsilon, c+\varepsilon)] \leq \tfrac{1}{2\varepsilon}\mathbb{E}[\max\{c-\varepsilon-X_t, 0\}].$$

Asymptotically, Doob's Upcrossing Inequality states that with $\varepsilon \to 0$,

$$\mathbb{E}[U(1-\varepsilon, 1+\varepsilon)] \in O\left(\tfrac{1}{\varepsilon}\right).$$

Again, we can use the construction of Theorem 6 to show that these asymptotics are tight: define the monotone decreasing function $f : \mathbb{N} \to [0, 1)$,

$$f(t) := \begin{cases} \frac{\delta}{2m}, & \text{if } t \leq m, \text{ and} \\ 0, & \text{otherwise.} \end{cases}$$

Then for $\delta = \frac{1}{2}$, Theorem 7 yields a martingale fulfilling the lower bound

$$\mathbb{E}[U(1-\tfrac{1}{4m}, 1+\tfrac{1}{4m})] \geq \frac{m}{2}$$

and Doob's Upcrossing Inequality gives the upper bound

$$\mathbb{E}[U(1-\tfrac{1}{4m}, 1+\tfrac{1}{4m})] \leq 2m,$$

which differs by a factor of 4.

The lower bound for the expected number of upcrossings given in Theorem 8 is a little looser than the upper bound given in Doob's Upcrossing Inequality. Closing this gap remains an open problem. We know by Theorem 9 that given a non-summable function f, the uniform probability for many $f(m)$-upcrossings goes to 0. However, this does not necessarily imply that expectation also tends to 0; low probability might be compensated for by high value. So for expectation there might be a lower bound larger than Theorem 7, an upper bound smaller than Doob's Upcrossing Inequality, or both.

If we drop the requirement that the rate of upcrossings to be uniform, Doob's Upcrossing Inequality is the best upper bound we can give [LH14].

6 Application to the MDL Principle

Let $\mathcal{M}$ be a countable set of probability measures on $(\Sigma^\omega, \mathcal{F}_\omega)$, called *environment class*. Let $K : \mathcal{M} \to [0, 1]$ be a function such that $\sum_{Q \in \mathcal{M}} 2^{-K(Q)} \leq 1$, called *complexity function on* $\mathcal{M}$. Following notation in [Hut09], we define for $u \in \Sigma^*$ the *minimal description length* model as

$$\mathrm{MDL}^u := \operatorname*{arg\,min}_{Q \in \mathcal{M}} \big\{ -\log Q(\Gamma_u) + K(Q) \big\}.$$

That is, $-\log Q(\Gamma_u)$ is the (arithmetic) code length of u given model Q, and $K(Q)$ is a complexity penalty for Q, also called *regularizer*. Given data $u \in \Sigma^*$, MDL^u is the measure $Q \in \mathcal{M}$ that minimizes the total code length of data and model.

The following corollary of Theorem 6 states that in some cases the limit $\lim_{t\to\infty} \mathrm{MDL}^{v_{1:t}}$ does not exist with high probability.

Corollary 13 (MDL May not Converge). *Let P be a probability measure on the measurable space $(\Sigma^\omega, \mathcal{F}_\omega)$. For any $\delta > 0$, there is a set of probability measures $\mathcal{M}$ containing P, a complexity function $K : \mathcal{M} \to [0, 1]$, and a measurable set $Z \in \mathcal{F}_\omega$ with $P(Z) \geq 1 - \delta$ such that for all $v \in Z$, the limit $\lim_{t\to\infty} \mathrm{MDL}^{v_{1:t}}$ does not exist.*

Proof. Fix some positive monotone decreasing summable function f (e.g., the one given in Theorem 8). Let $(X_t)_{t\in\mathbb{N}}$ be the P-martingale process from Theorem 6. By Theorem 5 there is a probability measure Q on $(\Sigma^\omega, \mathcal{F}_\omega)$ such that

$$X_t(v) = \frac{Q(\Gamma_{v_{1:t}})}{P(\Gamma_{v_{1:t}})}.$$

Choose $\mathcal{M} := \{P, Q\}$ with $K(P) := K(Q) := 1$. From the definition of MDL and Q it follows that

$$X_t(u) < 1 \iff Q(\Gamma_u) < P(\Gamma_u) \implies \mathrm{MDL}^u = P, \text{ and}$$
$$X_t(u) > 1 \iff Q(\Gamma_u) > P(\Gamma_u) \implies \mathrm{MDL}^u = Q.$$

For $Z := \bigcap_{m=1}^{\infty} E_{m,m}$ Theorem 6 yields

$$P(Z) = P[\forall m.\ U(1 - f(m), 1 + f(m)) \geq m] \geq 1 - \delta.$$

For each $v \in Z$, the measure $\mathrm{MDL}^{v_{1:t}}$ alternates between P and Q indefinitely, and thus its limit does not exist. □

Crucial to the proof of Theorem 13 is that not only does the process Q/P oscillate indefinitely, it oscillates around the constant $\exp(K(Q) - K(P)) = 1$. This implies that the MDL estimator may keep changing indefinitely, and thus it is inductively inconsistent.

7 Bounds on Mind Changes

Suppose we are testing a hypothesis $H \subseteq \Sigma^\omega$ on a stream of data $v \in \Sigma^\omega$. Let $P(H \mid \Gamma_{v_{1:t}})$ denote our belief in H at time $t \in \mathbb{N}$ after seeing the evidence $v_{1:t}$. By Bayes' rule,

$$P(H \mid \Gamma_{v_{1:t}}) = P(H) \frac{P(\Gamma_{v_{1:t}} \mid H)}{P(\Gamma_{v_{1:t}})} =: X_t(v).$$

Since X_t is a constant multiple of $P(\,\cdot \mid H)/P$ and $P(\,\cdot \mid H)$ is a probability measure on $(\Sigma^\omega, \mathcal{F}_\omega)$, the process $(X_t)_{t\in\mathbb{N}}$ is a P-martingale with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{N}}$ by Theorem 4. By definition, $(X_t)_{t\in\mathbb{N}}$ is bounded between 0 and 1. Let $\alpha > 0$. We are interested in the question how likely it is to often change one's mind about H by at least α, i.e., what is the probability for $X_t = P(H \mid \Gamma_{v_{1:t}})$ to decrease and subsequently increase m times by at least α. Formally, we define the stopping times $T'_{0,\nu}(v) := 0$,

$$T'_{2k+1,\nu}(v) := \inf\{t > T'_{2k,\nu}(v) \mid X_t(v) \leq X_{T'_{2k,\nu}(v)}(v) - \nu\alpha\},$$
$$T'_{2k+2,\nu}(v) := \inf\{t > T'_{2k+1,\nu}(v) \mid X_t(v) \geq X_{T'_{2k+1,\nu}(v)}(v) + \nu\alpha\},$$

and $T'_k := \min\{T'_{k,\nu} \mid \nu \in \{-1, +1\}\}$. (In Davis' notation, $X_{T'_{0,\nu}}, X_{T'_{1,\nu}}, \ldots$ is an α-alternating W-sequence for $\nu = 1$ and an α-alternating M-sequence for $\nu = -1$ [Dav13, Def. 4].) For any $t \in \mathbb{N}$, the random variable

$$A_t^X(\alpha)(v) := \sup\{k \geq 0 \mid T'_k(v) \leq t\},$$

is defined as the number of α-*alternations* up to time t. Let $A^X(\alpha) := \sup_{t\in\mathbb{N}} A_t^X(\alpha)$ denote the total number of α-alternations.

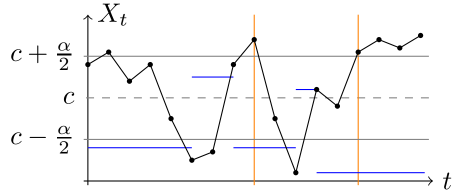

Fig. 2. This example process has two upcrossings between $c - \alpha/2$ and $c + \alpha/2$ (completed at the time steps of the vertical orange bars) and four α-alternations (completed when crossing the horizontal blue bars)

Setting $\alpha = 2\varepsilon$, the α-alternations differ from ε-upcrossings in three ways: first, for upcrossings, the process decreases below $c - \varepsilon$, then increases above $c + \varepsilon$, and then repeats. For alternations, the process may overshoot $c - \varepsilon$ or $c + \varepsilon$ and thus change the bar for the subsequent alternations, causing a 'drift' in the target bars over time. Second, for α-alternations the initial value of the martingale is relevant. Third, one upcrossing corresponds to two alternations, since one upcrossing always involves a preceding downcrossing. See Figure 2.

To apply our bounds for upcrossings on α-alternations, we use the following lemma by Davis. We reinterpret it as stating that every bounded martingale process $(X_t)_{t\in\mathbb{N}}$ can be modified into a martingale $(Y_t)_{t\in\mathbb{N}}$ such that the probability for many α-alternations is not decreased and the number of alternations equals the number of upcrossings plus the number of downcrossings [LH14].

Lemma 14 (Upcrossings and Alternations [Dav13, Lem. 9]). *Let $(X_t)_{t\in\mathbb{N}}$ be a martingale with $0 \leq X_t \leq 1$. There exists a martingale $(Y_t)_{t\in\mathbb{N}}$ with $0 \leq Y_t \leq 1$ and a constant $c \in (\alpha/2, 1-\alpha/2)$ such that for all $t \in \mathbb{N}$ and for all $k \in \mathbb{N}$,*

$$P[A_t^X(\alpha) \geq 2k] \leq P[A_t^Y(\alpha) \geq 2k] = P[U_t^Y(c-\alpha/2, c+\alpha/2) \geq k].$$

Theorem 15 (Upper Bound on Alternations). *For every martingale process $(X_t)_{t\in\mathbb{N}}$ with $0 \leq X_t \leq 1$,*

$$P[A(\alpha) \geq 2k] \leq \left(\frac{1-\alpha}{1+\alpha}\right)^k.$$

Proof. We apply Theorem 14 to $(X_t)_{t\in\mathbb{N}}$ and $(1-X_t)_{t\in\mathbb{N}}$ to get the processes $(Y_t)_{t\in\mathbb{N}}$ and $(Z_t)_{t\in\mathbb{N}}$. Dubins' Inequality yields

$$P[A_t^X(\alpha) \geq 2k] \leq P[U_t^Y(c_+ - \tfrac{\alpha}{2}, c_+ - \tfrac{\alpha}{2}) \geq k] \leq \left(\frac{c_+ - \frac{\alpha}{2}}{c_+ + \frac{\alpha}{2}}\right)^k =: g(c_+) \text{ and}$$

$$P[A_t^{1-X}(\alpha) \geq 2k] \leq P[U_t^Z(c_- - \tfrac{\alpha}{2}, c_- - \tfrac{\alpha}{2}) \geq k] \leq \left(\frac{c_- - \frac{\alpha}{2}}{c_- + \frac{\alpha}{2}}\right)^k = g(c_-)$$

for some $c_+, c_- \in (\alpha/2, 1-\alpha/2)$. Because Theorem 14 is symmetric for $(X_t)_{t\in\mathbb{N}}$ and $(1-X_t)_{t\in\mathbb{N}}$, we have $c_+ = 1-c_-$. Since $P[A_t^X(\alpha) \geq 2k] = P[A_t^{1-X}(\alpha) \geq 2k]$ by the definition of $A_t^X(\alpha)$, we have that both are less than $\min\{g(c_+), g(c_-)\} = \min\{g(c_+), g(1-c_+)\}$. This is maximized for $c_+ = c_- = 1/2$ because g is strictly monotone increasing for $c > \alpha/2$. Therefore

$$P[A_t^X(\alpha) \geq 2k] \leq \left(\frac{\frac{1}{2} - \frac{\alpha}{2}}{\frac{1}{2} + \frac{\alpha}{2}}\right)^k = \left(\frac{1-\alpha}{1+\alpha}\right)^k.$$

Since this bound is independent of t, it also holds for $P[A^X(\alpha) \geq 2k]$. □

The bound of Theorem 15 is the square root of the bound derived by Davis [Dav13, Thm. 10 & Thm. 11]:

$$P[A(\alpha) \geq 2k] \leq \left(\frac{1-\alpha}{1+\alpha}\right)^{2k} \tag{1}$$

This bound is tight [Dav13, Cor. 13].

Because $0 \leq X_t \leq 1$, the process $(1 - X_t)_{t \in \mathbb{N}}$ is also a nonnegative martingale, hence the same upper bounds apply to it. This explains why the result in Theorem 15 is worse than Davis' bound (1): Dubins' bound applies to all nonnegative martingales, while Davis' bound uses the fact that the process is bounded from below *and* above. For unbounded nonnegative martingales, downcrossings are 'free' in the sense that one can make a downcrossing almost surely successful (as done in the proof of Theorem 6). If we apply Dubins' bound to the process $(1 - X_t)_{t \in \mathbb{N}}$, we get the same probability bound for the downcrossings of $(X_t)_{t \in \mathbb{N}}$ (which are upcrossings of $(1 - X_t)_{t \in \mathbb{N}}$). Multiplying both bounds yields Davis' bound (1); however, we still require a formal argument why the upcrossing and downcrossing bounds are independent.

The following corollary to Theorem 15 derives an upper bound on the *expected* number of α-alternations.

Theorem 16 (Upper Bound on Expected Alternations). *For every martingale $(X_t)_{t \in \mathbb{N}}$ with $0 \leq X_t \leq 1$, the expectation $\mathbb{E}[A(\alpha)] \leq \frac{1}{\alpha}$.*

Proof. By Theorem 15 we have $P[A(\alpha) \geq 2k] \leq \left(\frac{1-\alpha}{1+\alpha}\right)^k$, and thus

$$\begin{aligned}
\mathbb{E}[A(\alpha)] &= \sum_{k=1}^{\infty} P[A(\alpha) \geq k] \\
&= P[A(\alpha) \geq 1] + \sum_{k=1}^{\infty} \big(P[A(\alpha) \geq 2k] + P[A(\alpha) \geq 2k+1]\big) \\
&\leq 1 + \sum_{k=1}^{\infty} 2P[A(\alpha) \geq 2k] \leq 1 + 2\sum_{k=1}^{\infty} \left(\frac{1-\alpha}{1+\alpha}\right)^k = \frac{1}{\alpha}. \qquad \square
\end{aligned}$$

We now apply the technical results of this section to the martingale process $X_t = P(\,\cdot \mid H)/P$, our belief in the hypothesis H as we observe data. The probability of changing our mind k times by at least α decreases exponentially with k (Theorem 15). Furthermore, the expected number of times we change our mind by at least α is bounded by $1/\alpha$ (Theorem 16). In other words, having to change one's mind a lot often is unlikely.

Because in this section we consider martingales that are bounded between 0 and 1, the lower bounds from Section 4 do not apply here. While for the martingales constructed in Theorem 6, the number of 2α-alternations and the number of α-up- and downcrossings coincide, these processes are not bounded. However, we can give a similar construction that is bounded between 0 and 1 and makes Davis' bound asymptotically tight.

8 Conclusion

We constructed an indefinitely oscillating martingale process from a summable function f. Theorem 6 and Theorem 7 give uniform lower bounds on the probability and expectation of the number of upcrossings of decreasing magnitude.

In Theorem 9 we proved the corresponding upper bound if the function f is not summable. In comparison, Doob's Upcrossing Inequality and Dubins' Inequality give upper bounds that are not uniform. In Section 5 we showed that for a certain summable function f, our martingale makes these bounds asymptotically tight as well.

Our investigation of indefinitely oscillating martingales was motivated by two applications. First, in Theorem 13 we showed that the minimum description length operator may not exist in the limit: for any probability measure P we can construct a probability measure Q such that Q/P oscillates forever around the specific constant that causes $\lim_{t\to\infty} \mathrm{MDL}^{v_{1:t}}$ to not converge.

Second, we derived bounds for the probability of changing one's mind about a hypothesis H when observing a stream of data $v \in \Sigma^\omega$. The probability $P(H \mid \Gamma_{v_{1:t}})$ is a martingale and in Theorem 15 we proved that the probability of changing the belief in H often by at least α decreases exponentially.

A question that remains open is whether there is a *uniform* upper bound on the *expected* number of upcrossings tighter than Doob's Upcrossing Inequality.

References

[Dav13] Davis, E.: Bounding changes in probability over time: It is unlikely that you will change your mind very much very often. Technical report (2013), `https://cs.nyu.edu/davise/papers/dither.pdf`

[Doo53] Doob, J.L.: Stochastic Processes. Wiley, New York (1953)

[Dur10] Durrett, R.: Probability: Theory and Examples, 4th edn. Cambridge University Press (2010)

[Grü07] Grünwald, P.D.: The Minimum Description Length Principle. The MIT Press, Cambridge (2007)

[Hut09] Hutter, M.: Discrete MDL predicts in total variation. In: Advances in Neural Information Processing Systems 22 (NIPS 2009), pp. 817–825. Curran Associates, Cambridge (2009)

[LH14] Leike, J., Hutter, M.: Indefinitely oscillating martingales. Technical report, The Australian National University (2014), `http://www.hutter1.net/publ/martoscx.pdf`

[LS05] Luo, W., Schulte, O.: Mind change efficient learning. In: Auer, P., Meir, R. (eds.) COLT 2005. LNCS (LNAI), vol. 3559, pp. 398–412. Springer, Heidelberg (2005)

[PH05] Poland, J., Hutter, M.: Asymptotics of discrete MDL for online prediction. IEEE Transactions on Information Theory 51(11), 3780–3795 (2005)

[Ris78] Rissanen, J.: Modeling by shortest data description. Automatica 14(5), 465–471 (1978)

[Wal05] Wallace, C.S.: Statistical and Inductive Inference by Minimum Message Length. Springer, Berlin (2005)

[WB68] Wallace, C.S., Boulton, D.M.: An information measure for classification. Computer Journal 11(2), 185–194 (1968)

[Xu12] Xu, W.: Martingale convergence theorems. Technical report (2012), `http://people.maths.ox.ac.uk/xu/Martingale_convergence.pdf`