Conditional Expectation and Martingales

Supplemental Material for Advanced Channel Coding
Henry D. Pfister

April 26th, 2014

1 Introduction

These notes are intended to introduce concentration inequalities for martingales with bounded increments. The final section also provides a gentle introduction to conditional expectation based on sigma fields. In contrast to providing a firm foundation for measure-theoretic probability, the primary goal is to introduce the language and intuition used in the study of martingales. The presentation is based on [1] but adds a few examples and details.

2 Background

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a countable sample space Ω. In this case, one can choose $\mathcal{F} = 2^{\Omega}$ and the probability of any event $A \in \mathcal{F}$ is given by the convergent sum

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega).$$

A random variable (r.v.) X is defined by a real function $X : \Omega \to \mathbb{R}$ of the random outcome $\omega \in \Omega$. The expected value of r.v. is denoted

$$\mathbb{E}[X] \triangleq \sum_{\omega \in \Omega} \mathbb{P}(\omega) X(\omega).$$

Let X, Y, Z be random variables on a common probability space. In this case, the conditional expectation can be seen both as the deterministic quantity

$$h(y) \triangleq \mathbb{E}\left[X|Y = y\right] = \sum_{\omega \in \Omega : Y(\omega) = y} \frac{\mathbb{P}\left(\omega\right)}{\mathbb{P}\left(\{\omega \in \Omega \,|\, Y = y\}\right)} X(\omega)$$

and as a new random variable

$$\mathbb{E}\left[X|Y\right] = h(Y).$$

Exercise 1. Use elementary probability theory to verify the following identities

$$\mathbb{E}\left[\mathbb{E}[X|Y]\right] = \mathbb{E}[X]$$
$$\mathbb{E}\left[\mathbb{E}[X|Y, Z]|Z\right] = \mathbb{E}[X|Z].$$

Example 2. Let $\Omega = \{1, 2, 3\}^2$ represent the outcomes of rolling two distinguishable 3-sided dice at once. Assume all outcomes are equiprobable and consider the r.v.s X, Y defined by $X\left((a, b)\right) = a$ and $Y\left((a, b)\right) = a + b$. Then, $Z = \mathbb{E}\left[X|Y\right]$ is a discrete r.v. with p.m.f. $\mathbb{P}\left(Z = z\right)$ defined by

$$\mathbb{P}\left(Z = z\right) = \begin{cases} \frac{1}{9} & \text{if } z = 1 \\ \frac{2}{9} & \text{if } z = 1.5 \\ \frac{3}{9} & \text{if } z = 2 \\ \frac{2}{9} & \text{if } z = 2.5 \\ \frac{1}{9} & \text{if } z = 3 \\ 0 & \text{otherwise.} \end{cases}$$

Definition 3. A r.v. X is called **integrable** if $\mathbb{E}[|X|] < \infty$. A sequence $(X_i)_{i=0}^{\infty}$ of r.v.s is called **uniformly integrable** if

$$\lim_{x\to\infty} \sup_i \mathbb{E}\left[|X_i| I_{\{|X|\geq x\}}\right] = 0,$$

where I_A is the indicator r.v. of the event $A \in \mathcal{F}$.

3 Martingales

Let $(X_i)_{i=0}^{\infty}$ and $(Z_i)_{i=0}^{\infty}$ be sequences of random variables (r.v.s) defined on a common probability space.

Definition 4. The sequence $(X_i)_{i=0}^{\infty}$ is called a **martingale** with respect to $(Z_i)_{i=0}^{\infty}$ if each X_i is integrable and, for $i \in \mathbb{N}$,

$$\mathbb{E}\left[X_i | Z_0, Z_1, \ldots Z_{i-1}\right] = X_{i-1}.$$

It is called a **supermartingale** w.r.t. $(Z_i)_{i=0}^{\infty}$ if $\mathbb{E}\left[X_i | Z_0, Z_1, \ldots Z_{i-1}\right] \leq X_{i-1}$ and a **submartingale** w.r.t. $(Z_i)_{i=0}^{\infty}$ if $\mathbb{E}\left[X_i | Z_0, Z_1, \ldots Z_{i-1}\right] \geq X_{i-1}$. If $X_i = Z_i$ for all i, then $(X_i)_{i=0}^{\infty}$ is simply called a **martingale** and

$$\mathbb{E}\left[X_i | X_0, X_1, \ldots X_{i-1}\right] = X_{i-1}.$$

Martingales were motivated by the idea of a gambler playing a fair game whose bet at time i can be any function of the past outcomes. Under this restriction, one finds that the expected wealth of the gambler satisfies

$$\mathbb{E}\left[X_i | X_0\right] = \mathbb{E}\left[\mathbb{E}\left[X_i | X_0, X_1, \ldots X_{i-1}\right] | X_0\right] = X_0$$

for all i. Taking the expected value on both sides of this equation shows that $\mathbb{E}\left[X_i\right] = \mathbb{E}\left[X_0\right]$.

The above results appear to contradict the well-known martingale betting strategy where a gambler bets \$1 initially and doubles her bet each time she loses. Eventually, she is guaranteed to win and, at that point, she will always make a \$1 profit. Thus, we have constructed a random stopping time T, which is chosen causally based on available information, such that $\mathbb{E}\left[X_T | X_0\right] = X_0 + 1$. This "paradox" was resolved by Doob's optional stopping time theorem [2, p. 261] which says that $\mathbb{E}\left[X_T\right] = \mathbb{E}\left[X_0\right]$ if: (i) $\mathbb{P}(T < \infty) = 1$, $\mathbb{E}\left[|X_T|\right] < \infty$, and $\lim_{n\to\infty} \mathbb{E}\left[X_n I_{\{T>n\}}\right] = 0$. In the above example, the first two conditions hold (e.g., T is geometric with mean 2 and $\mathbb{E}\left[|X_T|\right] = \mathbb{E}\left[|X_0 + 1|\right] < \infty$) but the third does not. Thus, the theorem does not apply.

Exercise 5. Let $S_n = Y_1 + Y_2 + \cdots + Y_n$ be the sum of n i.i.d. r.v. satisfying $\mathbb{E}\left[Y_i\right] = \mu < \infty$, $\mathbb{E}\left[(Y_i - \mu)^2\right] = \sigma^2 < \infty$, and $\phi(t) = \mathbb{E}\left[e^{tY_i}\right]$. Show that the following sequences are martingales:

1. $X_n = S_n - n\mu$
2. $X_n = (S_n - n\mu)^2 - n\sigma^2$
3. $X_n(t) = e^{tS_n - n\ln\phi(t)}$ for any fixed t such that $\phi(t) < \infty$.

Definition 6. Let $(X_i)_{i=0}^{\infty}$ be a martingale and define $Y_i = X_i - X_{i-1}$ for $i \in \mathbb{N}$. Then, Y_i is called a **martingale difference sequence**.

Exercise 7. Show that r.v.s in a martingale difference sequence are conditionally zero-mean (i.e., $\mathbb{E}\left[Y_i | X_0, \ldots, X_{i-1}\right] = 0$) and uncorrelated (i.e., $\mathbb{E}\left[Y_i Y_j | X_0, \ldots, X_{i-1}\right] = 0$ for $j < i$). Use this to show that, if $X_0 = 0$, then

$$\mathbb{E}\left[X_n^2\right] = \sum_{i=1}^{n} \mathbb{E}\left[Y_i^2\right] = \sum_{i=1}^{n} \mathbb{E}\left[(X_i - X_{i-1})^2\right].$$

Example 8 (Doob Martingale)**.** Let X be an integrable r.v. and $(Z_i)_{i=0}^{\infty}$ be an arbitrary sequence of random variables. Then, the sequence $X_i = \mathbb{E}\left[X | Z_0, Z_1, \ldots, Z_i\right]$, for $i = 0, 1, \ldots$, is automatically a martingale with respect to $(Z_i)_{i=0}^{\infty}$. One can verify this by observing that

$$\mathbb{E}\left[X_i | Z_0, Z_1, \ldots, Z_{i-1}\right] = \mathbb{E}\left[\mathbb{E}\left[X | Z_0, Z_1, \ldots, Z_i\right] | Z_0, Z_1, \ldots, Z_{i-1}\right] = \mathbb{E}\left[X | Z_0, Z_1, \ldots, Z_{i-1}\right] = X_{i-1}.$$

The sequence $(X_i)_{i=0}^{\infty}$ can be seen as increasingly accurate estimates of X which are constructed based on the observations $(Z_i)_{i=0}^{\infty}$. It can also be shown that the sequence $(X_i)_{i=0}^{\infty}$ is **uniformly integrable** .

Theorem 9 (Martingale Convergence)**.** *Let $(X_i)_{i=0}^{\infty}$ be a martingale satisfying* $\sup_i \mathbb{E}[|X_i|] < \infty$. *Then, the limit $X(\omega) = \lim_{n\to\infty} X_n(\omega)$ exists for almost all $\omega \in \Omega$. Moreover, $\mathbb{E}[|X|] < \infty$ and X_n converges to X almost surely. If, in addition, $(X_i)_{i=0}^{\infty}$ is a uniformly integrable martingale, then $\lim_{n\to\infty} \mathbb{E}[|X_i - X|] = 0$ and $X_i = \mathbb{E}[X|Z_0, Z_1, \ldots, Z_i]$. Note: the first result is used to prove that polar codes achieve capacity.*

Proof. See [2, p. 278]. □

4 Martingales with Bounded Differences

Martingales with bounded differences are particularly useful because they are easy to analyze and can be used to solve many problems in electrical engineering and computer science. They are often used to prove that a r.v. is tightly concentrated around its mean value. The interesting part is that the argument is independent of the mean value. The main result depends on the following lemma that was applied to i.i.d. sums by Hoeffding and to martingales with bounded differences by Azuma.

Lemma 10. *If the r.v. Z satisfies $|Z| \le c$ and $\mathbb{E}[Z|A] = 0$, then $\mathbb{E}\left[e^{\gamma Z}|A\right] \le e^{\gamma^2 c^2/2}$.*

Proof. Let $f(z) = e^{\gamma z}$ and observe that, since $f(z)$ is convex, it is upper bounded by the line segment connecting $(-c, f(-c))$ to $(c, f(c))$. This gives

$$f(z) \le g(z) = \frac{1}{2c}\left(e^{\gamma c} - e^{-\gamma c}\right) z + \frac{1}{2}\left(e^{\gamma c} + e^{-\gamma c}\right) = \frac{z}{c}\sinh(\gamma c) + \cosh(\gamma c).$$

From this, we determine that

$$\mathbb{E}\left[e^{\gamma Z}|A\right] \le \mathbb{E}[g(Z)|A] = \mathbb{E}[Z|A]\,\frac{1}{c}\sinh(\gamma c) + \cosh(\gamma c) = \cosh(\gamma c).$$

The final result follows from observing that $(2k)! = 2k(2k-1)\cdots 1 \ge 2^k k(k-1)\cdots 1$ and applying the upper bound

$$\cosh(x) = \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!} \le \sum_{k=0}^{\infty} \frac{x^{2k}}{k!2^k} \le \sum_{k=0}^{\infty} \frac{\left(\frac{x^2}{2}\right)^k}{k!} = e^{x^2/2}.$$

□

Theorem 11. *Let $X_i = \mathbb{E}[X|Z_0, Z_1, \ldots, Z_i]$ be a Doob martingale such that, for $i \in \mathbb{N}$,*

$$|X_i - X_{i-1}| \le c_i < \infty.$$

Then, for all $n \ge 1$ and any $\alpha > 0$, we have

$$\mathbb{P}\left(X_n - X_0 \ge \alpha\sqrt{n}\right) \le e^{-\frac{\alpha^2 n}{2\sum_{i=1}^n c_i^2}}$$

$$\mathbb{P}\left(X_0 - X_n \ge \alpha\sqrt{n}\right) \le e^{-\frac{\alpha^2 n}{2\sum_{i=1}^n c_i^2}}.$$

Proof. The idea is to estimate a Chernoff bound on $X_n - X_0$ using martingale properties. For any $\gamma \ge 0$, the Chernoff bound implies

$$\mathbb{P}\left(X_n - X_0 \ge \alpha\sqrt{n}\right) \le \frac{\mathbb{E}\left[e^{\gamma(X_n - X_0)}\right]}{e^{\gamma\alpha\sqrt{n}}}.$$

The expectation on the RHS can be upper bounded using

$$\begin{aligned}
\mathbb{E}\left[e^{\gamma(X_n - X_0)}\right] &= \mathbb{E}\left[e^{\gamma(X_{n-1} - X_0) + \gamma(X_n - X_{n-1})}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[e^{\gamma(X_{n-1} - X_0) + \gamma(X_n - X_{n-1})}|Z_0, Z_1, \ldots, Z_{n-1}\right]\right] && \text{(nested conditional } \mathbb{E}) \\
&= \mathbb{E}\left[e^{\gamma(X_{n-1} - X_0)}\mathbb{E}\left[e^{\gamma(X_n - X_{n-1})}|Z_0, Z_1, \ldots, Z_{n-1}\right]\right] && (X_{n-1}, X_0 \text{functions of } Z_0^{n-1}) \\
&\le \mathbb{E}\left[e^{\gamma(X_{n-1} - X_0)} e^{c_n^2\gamma^2/2}\right] && (\text{Lemma 10 via } \mathbb{E}\left[X_n - X_{n-1}|Z_0^{n-1}\right] = 0) \\
&\le e^{\frac{\gamma^2}{2}\sum_{i=1}^n c_i^2}. && \text{(by induction)}
\end{aligned}$$

The first stated result is obtained by using this bound and choosing $\gamma = \frac{\alpha\sqrt{n}}{\sum_i c_i^2}$. The second result follows from observing that, because

$$\mathbb{E}\left[e^{-\gamma Y_n} | Z_0, Z_1, \ldots, Z_{n-1}\right] \leq e^{c_n^2 \gamma^2/2},$$

one can also apply the same argument to $X_0 - X_n$. □

Theorem 12 (McDiarmid)**.** *Let $Z_1, Z_2, \ldots, Z_n$ be independent random variables taking values in $\mathcal{Z}$ and $f : \mathcal{Z}^n \to \mathbb{R}$ be a real function. Then,*

$$\mathbb{P}\left(f(Z_1, \ldots, Z_n) - \mathbb{E}\left[f(Z_1, \ldots, Z_n)\right] \geq \alpha\sqrt{n}\right) \leq e^{-\frac{\alpha^2 n}{2\sum_{i=1}^n c_i^2}},$$

where the Lipschitz constant (w.r.t. Hamming distance) associated with the i-th coordinate is

$$c_i = \sup_{z_1,\ldots,z_n,z_i' \in \mathcal{Z}} \left|f(z_1, \ldots, z_n) - f(z_1, \ldots, z_{i-1}, z_i', z_{i+1}, \ldots, z_n)\right|.$$

Proof. The proof is written for discrete r.v.s but can easily be extended to the general case. Let $X = f(Z_1, \ldots, Z_n)$ and $X_i = \mathbb{E}\left[X | Z_1, Z_2, \ldots, Z_i\right]$ define a Doob martingale with respect to $(Z_i)_{i=1}^n$. Since $Z_1, Z_2, \ldots, Z_n$ are independent r.v., we observe that

$$\mathbb{E}\left[f(Z_1, \ldots, Z_n) \,|\, Z_1, \ldots, Z_{i-1}\right] = \sum_{z_i' \in \mathcal{Z}} \mathbb{P}(Z_i = z_i') \mathbb{E}\left[f(Z_1, \ldots, Z_{i-1}, z_i', Z_{i+1}, \ldots, Z_n) \,|\, Z_1, \ldots, Z_i\right].$$

Using this, we can write

$$\begin{aligned}
|X_i - X_{i-1}| &= \left|\mathbb{E}\left[f(Z_1, \ldots, Z_n) \,|\, Z_1, \ldots, Z_i\right] - \mathbb{E}\left[f(Z_1, \ldots, Z_n) \,|\, Z_1, \ldots, Z_{i-1}\right]\right| \\
&= \left|\mathbb{E}\left[f(Z_1, \ldots, Z_n) \,|\, Z_1, \ldots, Z_i\right] - \sum_{z_i' \in \mathcal{Z}} \mathbb{P}(Z_i = z_i') \mathbb{E}\left[f(Z_1, \ldots, Z_{i-1}, z_i', Z_{i+1}, \ldots, Z_n) \,|\, Z_1, \ldots, Z_i\right]\right| \\
&= \left|\sum_{z_i' \in \mathcal{Z}} \mathbb{P}(Z_i = z_i') \mathbb{E}\left[f(Z_1, \ldots, Z_n) - f(Z_1, \ldots, Z_{i-1}, z_i', Z_{i+1}, \ldots, Z_n) \,|\, Z_1, \ldots, Z_i\right]\right| \\
&\leq \sum_{z_i' \in \mathcal{Z}} \mathbb{P}(Z_i = z_i') \mathbb{E}\left[\left|f(Z_1, \ldots, Z_n) - f(Z_1, \ldots, Z_{i-1}, z_i', Z_{i+1}, \ldots, Z_n)\right| \,|\, Z_1, \ldots, Z_i\right] \\
&\leq \sup_{z_1,\ldots,z_n,z_i' \in \mathcal{Z}} \left|f(z_1, \ldots, z_n) - f(z_1, \ldots, z_{i-1}, z_i', z_{i+1}, \ldots, z_n)\right|.
\end{aligned}$$

Since $X_0 = \mathbb{E}\left[f(Z_1, \ldots, Z_n)\right]$ and $X_n = f(Z_1, \ldots, Z_n)$, we can apply Theorem 11 to obtain the stated result.

□

Definition 13. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space for the set of graphs with n vertices. Let G be a random graph and suppose that its vertices are labeled $1, 2, \ldots, n$ and exposed sequentially in that order. Let the random variable Z_i define all the edges emananting from vertex i. Then, for any function $f : \Omega \to \mathbb{R}$, the Doob martingale $X_i = \mathbb{E}\left[f(G) | Z_1, Z_2, \ldots, Z_i\right]$ is called the **vertex exposure martingale** and satisfies $X_0 = \mathbb{E}\left[f(G)\right]$ and $X_n = f(G)$.

At each step, one new vertex is observed along with all edges connecting that vertex to previously exposed vertices. Let E_{jk}, for $1 \leq j < k \leq n$, be the indicator r.v. of an edge connecting vertex j to vertex k. The observation process can be seen to expose subsets of these indicator functions at each step. In particular, we have $Z_i = \{E_{ji}\}_{1 \leq j < i}$.

Example 14. The chromatic number $\chi(G)$ of a graph G is the minimum number of colors such that, when all vertices are assigned colors, each pair of adjacent vertices can be assigned different colors. Let $G(n, p)$ be the Erdos-Renyi ensemble of random graphs with n vertices where each of the $\binom{n}{2}$

possible edges are chosen independently with probability p. Let G be a random graph from $G(n,p)$ and $X_i = \mathbb{E}\left[\chi(G)|Z_1, Z_2, \ldots, Z_i\right]$ be the vertex exposure martingale for the chromatic number of G.

For a graph G, let G_i be the subgraph generated by the first i vertices of G and observe that X_i is the weighted average of the chromatic number over all possible extensions of G_i. Likewise, X_{i-1} is the weighted average of the chromatic number over all possible extensions of G_{i-1}. In this average, one can group together all terms that have the same edge connections between vertex i and vertices $< i$. Doing this, shows that $X_i - X_{i-1}$ is upper bounded by the maximum change in the chromatic number associated with arbitrary changes to edges between vertex i and vertices $< i$, while keeping all other edges fixed.

Fortunately, adding one vertex (along with all of its edges) to a graph can increase the chromatic number by at most one (e.g., it can always be assigned the new color). Since the chromatic number is unchanged by deleting a vertex and adding it back again, this also implies that deleting a vertex can reduce it by at most one. This implies that $|X_i - X_{i-1}| \leq 1$ and allows us to apply Theorem 11 to see that

$$\mathbb{P}\left(|\chi(G) - \mathbb{E}\left[\chi(G)\right]| \geq \alpha\sqrt{n}\right) \leq 2e^{-\alpha^2/2}.$$

Exercise 15. Does the previous analysis of the chromatic number apply to an arbitrary distribution over graphs? If so, explain. If not, give a counterexample.

Definition 16. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space for the set of graphs with n vertices. Let G be a random graph and suppose that its edges are exposed sequentially in a fixed order. Let $e = \binom{n}{2}$ and define E_i, for $i = 1, \ldots, e$, to be an indicator r.v. for the existence of the i-th exposed edge. Then, for any function $f : \Omega \to \mathbb{R}$, the Doob martingale $X_i = \mathbb{E}\left[f(G)|E_1, E_2, \ldots, E_i\right]$ is called the **edge exposure martingale** and satisfies $X_0 = \mathbb{E}\left[f(G)\right]$ and $X_e = f(G)$.

Exercise 17. Repeat the previous analysis of the chromatic number using an edge exposure martingale and the fact that adding/deleting one edge changes the chromatic number by at most one. Is the bound better or worse than the vertex exposure martingale?

Now, we extend this approach to handle functions of permutations. Let S_n be the symmetric group on n elements and, for $\pi, \sigma \in S_n$, let $d(\pi, \sigma)$ be the minimum number of transpositions required to transform π into σ (or vice-versa). It turns out that (S_n, d) forms a metric space.

Theorem 18. *Let $f : S_n \to \mathbb{R}$ be a function that satisfies $|f(\pi) - f(\sigma)| \leq c\, d(\pi, \sigma)$ for all $\pi, \sigma \in S_n$ (i.e., f is Lipschitz-c w.r.t. d). If $\Pi \in S_n$ is a uniform random permutation on n elements, then*

$$\mathbb{P}\left(f(\Pi) - \mathbb{E}\left[f(\Pi)\right] \geq \alpha\sqrt{n}\right) \leq e^{-\alpha^2/(2c^2)}.$$

Proof. Let Π be a random permutation with values $\Pi_1, \Pi_2, \ldots, \Pi_n$. Then, $X_j = \mathbb{E}\left[f(\Pi)\,|\,\Pi_1, \ldots, \Pi_j\right]$ defines a Doob martingale with respect to $(\Pi_i)_{i=1}^n$. The key observation we need is that

$$\mathbb{E}\left[f(\Pi)\,|\,\Pi_1, \ldots, \Pi_{j-1}\right] = \sum_{k=j}^{n} \frac{1}{n-j+1} \mathbb{E}\left[f\left((jk)\Pi\right)\,|\,\Pi_1, \ldots, \Pi_j\right],$$

where the transposition (jk) swaps the the j-th and k-th elements of Π. Using this, we can bound the increments with

$$\begin{aligned}
|X_j - X_{j-1}| &= \left|\mathbb{E}\left[f(\Pi)\,|\,\Pi_1, \ldots, \Pi_j\right] - \mathbb{E}\left[f(\Pi)\,|\,\Pi_1, \ldots, \Pi_{j-1}\right]\right| \\
&= \left|\mathbb{E}\left[f(\Pi)\,|\,\Pi_1, \ldots, \Pi_j\right] - \sum_{k=j}^{n} \frac{1}{n-j+1} \mathbb{E}\left[f\left((jk)\Pi\right)\,|\,\Pi_1, \ldots, \Pi_j\right]\right| \\
&= \left|\sum_{k=j}^{n} \frac{1}{n-j+1} \mathbb{E}\left[f(\Pi) - f\left((jk)\Pi\right)\,|\,\Pi_1, \ldots, \Pi_j\right]\right| \\
&\leq \sum_{k=j}^{n} \frac{1}{n-j+1} \mathbb{E}\left[\left|f(\Pi) - f\left((jk)\Pi\right)\right|\,|\,\Pi_1, \ldots, \Pi_j\right] \\
&\leq c.
\end{aligned}$$

Since $X_0 = \mathbb{E}\left[f(\Pi)\right]$ and $X_n = f(\Pi)$, we can apply Theorem 11 to obtain the stated result. □

Now, we use the two previous theorems to prove the concentration theorem for message-passing decoding of irregular LDPC codes. The only decoder property used is that the output message from any bit node depends only on the received values and the graph structure in a "small" local neighborhood of the bit node.

Corollary 19. *Consider a random code drawn from the LDPC(n, λ, ρ) ensemble, transmitted over a memoryless channel, and let $X_1, \ldots, X_n$ be the bit-node output messages after ℓ iterations of message-passing decoding. Let $h : \mathcal{M} \to \mathbb{R}$ satisfy $\sup_{x,x' \in \mathcal{M}} |h(x) - h(x')| \leq 1$. Then, we have*

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} h(X_i) - \mathbb{E}\left[\sum_{i=1}^{n} h(X_i)\right]\right| \geq \alpha\sqrt{n}\right) \leq 2e^{-\alpha^2/(2c^2)} + 2e^{-\alpha^2 d/(8c^2)},$$

where $c = \mathtt{r}_{\max}(\mathtt{r}_{\max}\,\mathtt{l}_{\max})^{\ell}$ and $d = \left(\int_0^1 \lambda(x)dx\right)^{-1}$.

Proof. First, we observe that the code has n bits and $m = dn$ edges. Therefore, the r.v. X_i is given by a deterministic function $g_i : \mathcal{Y}^n \times S_m \to \mathcal{M}$ of the received vector Y^n and the permutation Π that defines the code. This allows us to define

$$f(Y^n, \Pi) \triangleq \sum_{i=1}^{n} h(X_i) = \sum_{i=1}^{n} h\left(g_i(Y^n, \Pi)\right).$$

The proof proceeds in two steps and consists mainly of verifying that $f(Y^n, \Pi)$ is Lipschitz in Y^n w.r.t. to Hamming distance and Lipschitz in Π w.r.t. transposition distance. Next, we show that, for a fixed Π, H is concentrated around its average over Y^n. Finally, we show that its average over Π is concentrated around its overall expectation.

Let $\mathcal{N}_i^{(\ell)}$ be the subgraph generated by the check nodes in depth-ℓ neighborhood of the i-th bit and observe that the output message of the i-th bit depends only on $\mathcal{N}_i^{(\ell)}$ and the received values of bit nodes in $\mathcal{N}_i^{(\ell)}$. Since the maximum bit and check degrees are given by $\mathtt{l}_{\max}$ and $\mathtt{r}_{\max}$, the maximum number of check nodes in $\mathcal{N}_i^{(\ell)}$ is given by

$$\begin{aligned} B &= \mathtt{l}_{\max} \sum_{i=0}^{\ell-1} \left((\mathtt{l}_{\max} - 1)(\mathtt{r}_{\max} - 1)\right)^{\ell} \\ &= \mathtt{l}_{\max} \frac{\left((\mathtt{l}_{\max} - 1)(\mathtt{r}_{\max} - 1)\right)^{\ell} - 1}{(\mathtt{l}_{\max} - 1)(\mathtt{r}_{\max} - 1) - 1} \\ &\leq \left(\mathtt{r}_{\max}\mathtt{l}_{\max}\right)^{\ell}. \end{aligned}$$

Since each check node has at most $\mathtt{r}_{\max}$ edges, the maximum number of bit nodes (and edges) in $\mathcal{N}_i^{(\ell)}$ is upper bounded by $A = \mathtt{r}_{\max}B$.

Since bit j is in $\mathcal{N}_i^{(\ell)}$ iff bit i is in $\mathcal{N}_j^{(\ell)}$, it follows that changes in the i-th received value affect at most A output values. Therefore, for any $\pi \in S_m$, the maximum change in $f(Y^n, \pi)$ is upper bounded by the product of the number of X_i's that can change (i.e., A) with the maximum change in $h(X_i)$ (i.e., 1). The resulting upper bound on the Lipschitz constant is $c = A$ and Theorem 12 allows us to conclude that

$$\mathbb{P}\left(f(Y^n, \pi) - \mathbb{E}\left[f(Y^n, \pi)\right] \geq \alpha\sqrt{n}\right) \leq 2e^{-\alpha^2/(2c^2)}.$$

Swapping the endpoints of any two edges in the graph corresponds to transposing two elements of π in $f(y^n, \pi)$. Since this only affects the i-th output message if one of those edges is in $\mathcal{N}_i^{(\ell)}$, it follows that the value of $f(y^n, \pi)$ changes by at most $b = 2A = 2c$ (i.e., at most $2A$ values change by at most

1). Since this bound holds for all y^n, is also holds for any average over y^n. Therefore, we can apply Theorem 18 to $\tilde{f}(\pi) = \mathbb{E}\left[f(Y^n,\pi)\right]$ and conclude that

$$\mathbb{P}\left(\tilde{f}(\Pi) - \mathbb{E}\left[\tilde{f}(\Pi)\right] \geq \alpha\sqrt{n}\right) \leq 2e^{-\alpha^2 \frac{m}{n}/(2b^2)} = 2e^{-\alpha^2 d/(8c^2)}.$$

Combining these two bounds gives the stated result. □

5 Sigma Fields and Probability

Definition 20. A **σ-field** $(\Omega,\mathcal{F})$ consists of a set Ω and a collection of subsets $\mathcal{F} \subseteq 2^{\Omega}$ that satisfies

1. $\emptyset \in \mathcal{F}$
2. $A \in \mathcal{F}$ implies $A^c \in \mathcal{F}$
3. $A_i \in \mathcal{F}$ for $i \in \mathbb{N}$ implies $\cup_{i\in\mathbb{N}} A_i \in \mathcal{F}$

Remark 21. In this note, we restrict our attention to the case that Ω is a complete separable metric space. For a finite sample space Ω, it is often easiest to use the maximal σ-field $\mathcal{F} = 2^{\Omega}$. For infinite sample spaces, a common choice is the Borel σ-field $\mathcal{F} = \mathcal{B}(\Omega)$, which is the σ-field generated by all open sets in the topology generated by the metric.

Definition 22. Let Ω be a set and $\mathcal{A} \subseteq 2^{\Omega}$ be a collection of subsets. The σ-field **generated** by $\mathcal{A}$ is denoted $\sigma(\mathcal{A})$ and is defined to be the intersection of all σ-fields that contain $\mathcal{A}$ (i.e., the smallest σ-field that contains $\mathcal{A}$). Existence and uniqueness follows from the intersection in the definition.

Example 23. Consider the sample space $\Omega = \{0,1\}^2$ associated with two Bernoulli trials. The σ-field $\mathcal{F}_1 = \{\emptyset, \{00,01\}, \{11,10\}, \{00,01,10,11\}\}$ is generated by $\mathcal{A} = \{\{00,01\}\}$ and only distinguishes between the outcomes of the first trial. Likewise, $\mathcal{F}_2 = \{\emptyset, \{00,10\}, \{11,01\}, \{00,01,10,11\}\}$ only resolves outcomes of the second trial.

Definition 24. A **probability space** $(\Omega,\mathcal{F},\mathbb{P})$ consists of a σ-field $(\Omega,\mathcal{F})$ and a probability law $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ that satisfies

1. $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{F}$
2. $\mathbb{P}(\Omega) = 1$
3. If $A_i \in \mathcal{F}$, for $i \in \mathbb{N}$, is a collection of pairwise disjoint sets (i.e., $A_i \cap A_j = \emptyset$ for $i \neq j$), then
$$\mathbb{P}\left(\cup_{i\in\mathbb{N}} A_i\right) = \sum_{i\in\mathbb{N}} \mathbb{P}\left(A_i\right).$$

Example 25. Continuing the example, one observes that any probability law on $\mathcal{F}_1$ is completely determined by the success probability $\mathbb{P}\left(\{11,10\}\right)$ of the first trial. Likewise, any probability law on $\mathcal{F}_2$ is completely determined by $\mathbb{P}\left(\{11,01\}\right)$.

Definition 26. A function $X : \Omega \to \mathbb{R}$ is **measurable** w.r.t. the σ-field $\mathcal{F}$ (or $\mathcal{F}$-measurable) if the set $\{\omega \in \Omega \,|\, X(\omega) \leq a\}$ is in $\mathcal{F}$ for all $a \in \mathbb{R}$. Furthermore, we let $\mathcal{F}_X = \sigma(X)$ denote the σ-field generated by the aforementioned collection of sets.

Remark 27. There is a one-to-one correspondence between measurable functions and random variables.

Example 28. Continuing the previous example, we let $X_1 : \Omega \to \mathbb{R}$ be the function defined by $X_1\left((a,b)\right) = a$. Then, X_1 is $\mathcal{F}_1$-measurable because $\{\omega \in \Omega \,|\, X_1(\omega) \leq 0\} = \{00,01\} \in \mathcal{F}_1$ and $\{\omega \in \Omega \,|\, X_1(\omega) \leq 1\} = \{00,01,10,11\} \in \mathcal{F}_1$. But, X_1 is not $\mathcal{F}_2$-measurable because $\{00,01\} \notin \mathcal{F}_2$. One also finds the reverse is true for $X_2\left((a,b)\right) = b$.

Proposition 29. *Let X, Y be random variables (i.e., measurable functions) defined on $(\Omega, \mathcal{F}, \mathbb{P})$. If Y is $\mathcal{F}_X$-measurable, then Y is a deterministic function of X.*

Proof. We prove the contrapositive. If Y is not a deterministic function of X, then there is an $A \subseteq \Omega$ such that $X(A) = a$ and $Y(A)$ is not a singleton. For any $y \in (\inf Y(A), \sup Y(A))$, one finds that $\{\omega \in \Omega \,|\, Y(\omega) \leq y\} \subset \{\omega \in \Omega \,|\, X(\omega) = a\}$. But, no proper subset of $\{\omega \in \Omega \,|\, X(\omega) = a\}$ is in $\mathcal{F}_X$. This implies that Y is not $\mathcal{F}_X$-measurable. □

Definition 30. For a σ-algebra $(\Omega, \mathcal{F})$, let the **atom** containing ω be $\mathcal{A}_\omega(\mathcal{F}) = \cap_{A \in \mathcal{F}: \omega \in A} A$ or the smallest element of $\mathcal{F}$ that contains ω. The collection of all atoms in $\mathcal{F}$ is given by $\{A_\omega(\mathcal{F}) \,|\, \omega \in \Omega\}$.

Example 31. Continuing the previous example, we observe that, for $\mathcal{F}_1$, the collection of atoms is given by $\{\{00, 01\}, \{11, 10\}\}$.

Proposition 32. *Let X, Y be random variables (i.e., measurable functions) defined on $(\Omega, \mathcal{F}, \mathbb{P})$. If Y is $\mathcal{F}_X$-measurable, then Y is constant on all atoms of $\mathcal{F}_X$.*

Proof. The proof of the previous Proposition can be easily modified to handle this case. □

Example 33. Continuing the previous example, we check that X_1 is constant on the atoms of $\mathcal{F}_1$, since it is $\mathcal{F}_1$-measurable. For the first atom $\{00, 01\}$, we verify that $X_1(00) = X_1(01) = 0$. Checking the second atom $\{00, 01\}$ is left to the reader.

Remark 34. While atoms are well-defined for both countable and uncountable sample spaces, they are much more useful in countable sample spaces. In the countable case, every element in the σ-algebra can be constructed as a countable union of atoms. For uncountable sample spaces, the atoms may be the points (e.g., in the Borel σ-field) but general elements on the σ-field cannot be generated by countable unions of atoms.

Example 35. Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\Omega = [0, 1]^2$, where $\mathcal{F} = \mathcal{B}\left([0, 1]^2\right)$ is the Borel σ-algebra generated by the standard open sets on this space. The atoms of this σ-algebra are given by the points of Ω because each point can be seen as the countable intersection of open sets containing that point. Let $X((a, b)) = a$ be a r.v. on this probability space and consider the atoms of $\mathcal{F}_X = \sigma(X)$. In this case, one finds that $\mathcal{A}_{(a,b)}(\mathcal{F}_X)$ is the subspace $\{(a', b) \in \Omega \,|\, a' = a\}$. In other words, knowledge of X removes all uncertainty in one dimension but provides no information about the other.

6 Conditional Expectation and Sigma Fields

If the sample space Ω is countable, then the conditional expectation can be seen both as the deterministic quantity

$$h(y) \triangleq \mathbb{E}[X|Y = y] = \sum_{\omega \in \Omega : Y(\omega) = y} \frac{\mathbb{P}(\omega)}{\mathbb{P}(\{\omega \in \Omega \,|\, Y = y\})} X(\omega)$$

and the random variable

$$\mathbb{E}[X|Y] = h(Y).$$

When the sample space is uncountable, the sum in $h(y)$ becomes an integral, but the quantity $\mathbb{P}(\{Y = y\}) = \mathbb{P}(\{\omega \in \Omega \,|\, Y = y\})$ may be zero. For this reason, the function $h(y)$ may not be well-defined and the following definition is used.

Definition 36. Let X, Y be random variables (i.e., measurable functions) defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and let X be integrable. The **conditional expectation** $Z = \mathbb{E}[X|Y]$ is defined to be any random variable (i.e., measurable function) that satisfies

1. Z is $\mathcal{F}_Y$-measurable
2. For all $\Lambda \in \mathcal{F}_Y$, we have $\int_\Lambda Z(\omega)\, d\mathbb{P} = \int_\Lambda X(\omega)\, d\mathbb{P}$

Remark 37. One issue is that the conditional expectation is, strictly speaking, not unique. Any two r.v. satisfying the above conditions (for a fixed X, Y) must be equal almost everywhere (a.e.) but can differ on sets of measure zero. Another problem is existence. While the conditional expectation always exists, proving this requires some work and is neglected in these notes.

It can also be useful to consider conditional expectations w.r.t. to σ-fields.

Definition 38. Let X be an integrable random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$ be a subfield of $\mathcal{F}$. The **conditional expectation** $Z = \mathbb{E}\left[X|\mathcal{G}\right]$ is defined to be any random variable (i.e., measurable function) that satisfies

1. Z is $\mathcal{G}$-measurable
2. For all $\Lambda \in \mathcal{G}$, we have $\int_\Lambda Z(\omega)\, d\mathbb{P} = \int_\Lambda X(\omega)\, d\mathbb{P}$

Remark 39. One should notice that the only difference from the previous definition is that $\mathcal{F}_Y$ has been replaced by $\mathcal{G}$. This implies that $\mathbb{E}\left[X|Y\right] = \mathbb{E}\left[X|\mathcal{G}\right]$ if $\mathcal{G}$ is chosen to be the σ-field generated by Y (i.e., $\mathcal{G} = \sigma(Y) = \mathcal{F}_Y$). Likewise, this implies that $\mathbb{E}\left[X|\mathcal{G}\right] = \mathbb{E}\left[X|Y\right]$ if Y is chosen to be a r.v. that takes distinct values on all atoms of $\mathcal{G}$.

Example 40. Continuing the previous example, we let $\mathcal{G} = \sigma(Y)$ be the σ-field generated by Y and observe also that $\mathcal{G}$ is the σ-field generated by the atoms $\mathcal{A} = \{\{11\}, \{12, 21\}, \{13, 22, 31\}, \{23, 32\}, \{33\}\}$. It is also worth observing that $Z = E\left[X|\mathcal{G}\right]$ is constant on all atoms of $\mathcal{G}$.

7 Martingales and Filters

Definition 41. A **filter** w.r.t. a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is an increasing sequence $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}$ of σ-fields such that $(\Omega, \mathcal{F}_i, \mathbb{P})$, for $i = 0, 1, \ldots$, is a probability space. It is assumed that $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and, by definition, the σ-fields are ordered by refinement in the sense that $A \in \mathcal{F}_i$ implies $A \in \mathcal{F}_j$ for $i \leq j$.

Remark 42. One should think about the elements in $\mathcal{F}_i$ as summaries of the history of the random process $X_0, X_1, \ldots$ up to time i.

Proposition 43. *Let X be a r.v. and $(\mathcal{F}_i)_{i=0}^{\infty}$ be a filter on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, we have*

1. $\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{F}_i\right]|\mathcal{F}_j\right] = \mathbb{E}\left[X|\mathcal{F}_{\min\{i,j\}}\right]$ *for* $i, j \geq 0$
2. *If Y is $\mathcal{F}_i$-measurable, then* $\mathbb{E}\left[XY|\mathcal{F}_i\right] = \mathbb{E}\left[X|\mathcal{F}_i\right] Y$

Proof. TBD □

Definition 44. Let $(\mathcal{F}_i)_{i=0}^{\infty}$ be a filter on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $(X_i)_{i=0}^{\infty}$ be a sequence of integrable r.v.s such that X_i is $\mathcal{F}_i$-measurable. We say that $(X_i)_{i=0}^{\infty}$ $X_0, X_1, \ldots$ is a **martingale** w.r.t. $(\mathcal{F}_i)_{i=0}^{\infty}$ if, for $i \in \mathbb{N}$,

$$\mathbb{E}\left[X_i|\mathcal{F}_{i-1}\right] = X_{i-1}.$$

It is called a **supermartingale** if $\mathbb{E}\left[X_i|\mathcal{F}_{i-1}\right] \leq X_{i-1}$ and a **submartingale** if $\mathbb{E}\left[X_i|\mathcal{F}_{i-1}\right] \geq X_{i-1}$.

Remark 45. This reduces to the simpler definition of a martingale by choosing $\mathcal{F}_i = \sigma(X_0, \ldots, X_i)$ (i.e., the minimal filter that makes each X_i $\mathcal{F}_i$-measurable) so that $\mathbb{E}\left[X_i|\mathcal{F}_{i-1}\right] = \mathbb{E}\left[X_i|X_0, X_1, \ldots X_{i-1}\right]$.

Definition 46. Let $(X_i)_{i=0}^{\infty}$ be a martingale w.r.t. $(\mathcal{F}_i)_{i=0}^{\infty}$ and define, for $i = 1, 2, \ldots$, $Y_i = X_i - X_{i-1}$. Then, Y_i is called a **martingale difference sequence**.

Example 47 (Doob Martingale)**.** Let X be an integrable r.v. and $(\mathcal{F}_i)_{i=0}^{\infty}$ be a filter on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, the sequence of r.v.s $X_i = \mathbb{E}\left[X|\mathcal{F}_i\right]$ is automatically a uniformly integrable martingale. One can verify the martingale property by observing that

$$\mathbb{E}\left[X_i|\mathcal{F}_{i-1}\right] = \mathbb{E}\left[\mathbb{E}\left[X|\mathcal{F}_i\right]|\mathcal{F}_{i-1}\right] = \mathbb{E}\left[X|\mathcal{F}_{i-1}\right] = X_{i-1}.$$

The sequence of r.v.s can be seen as increasingly accurate estimates of X that are generated by the increasingly refined knowledge of X represented by the filter's increasing σ-fields.

Theorem 48 (Doob's Maximal Inequality)**.** *Let $(X_i)_{i=0}^{\infty}$ be a non-negative submartingale w.r.t. $(\mathcal{F}_i)_{i=0}^{\infty}$. Then, for any $p \geq 1$ and $\lambda > 0$, we have*

$$\mathbb{P}\left(\max_{i\in\{0,\ldots,n\}} X_i \geq \lambda\right) \leq \frac{\mathbb{E}\left[X_n^p\right]}{\lambda^p}.$$

Proof. For $p \geq 1$, if $(X_i)_{i=0}^{\infty}$ is a non-negative submartingale, then $(X_i^p)_{i=0}^{\infty}$ is too. Applying the maximal inequality to $(X_i^p)_{i=0}^{\infty}$ gives the stated result [2, p. 280]. □

Example 49. Let $(X_i)_{i=0}^{\infty}$ be a martingale w.r.t. $(\mathcal{F}_i)_{i=0}^{\infty}$. Then, for any $q \geq 1$, $Y_i = |X_i|^q$ is a non-negative submartingale. Thus, for any $p \geq 1$, we have

$$\mathbb{P}\left(\sup_{i\in\{0,\ldots,n\}} |X_i|^q \geq \lambda\right) \leq \frac{\mathbb{E}\left[|X_n|^{qp}\right]}{\lambda^p}.$$

Theorem 50 (Martingale Convergence)**.** *Let $(X_i)_{i=0}^{\infty}$ be a martingale w.r.t. $(\mathcal{F}_i)_{i=0}^{\infty}$ satisfying $\sup_{i\in\mathbb{N}} \mathbb{E}\left[|X_i|\right] < \infty$. Then, the limit $X(\omega) = \lim_{n\to\infty} X_n(\omega)$ exists for almost all $\omega \in \Omega$ and X is measurable w.r.t. $\mathcal{F}_\infty = \sigma\left(\cup_i \mathcal{F}_i\right)$. Moreover, $\mathbb{E}\left[|X|\right] < \infty$ and X_n converges to X almost surely. If, in addition, $(X_i)_{i=0}^{\infty}$ is a uniformly integrable martingale, then $\lim_{n\to\infty} \mathbb{E}\left[|X_i - X|\right] = 0$ and $X_i = \mathbb{E}\left[X|\mathcal{F}_i\right]$.*

Proof. See [2, p. 309]. □

References

[1] T. J. Richardson and R. L. Urbanke, *Modern Coding Theory.* New York, NY: Cambridge University Press, 2008.

[2] S. Karlin and H. M. Taylor, "A first course in stochastic processes," 1975.