On Concentration of Martingales and Applications in Information Theory, Communication & Coding

Igal Sason

Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel

ETH, Zurich, Switzerland
August 22–23, 2012.

Concentration of Measures

- Concentration of measures is a central issue in probability theory, and it is strongly related to information theory and coding.
- Roughly speaking, the concentration of measure phenomenon can be stated in the following simple way: "A random variable that depends in a smooth way on many independent random variables (but not too much on any of them) is essentially constant" (Talagrand, 1996).

Concentration of Measures

- Concentration of measures is a central issue in probability theory, and it is strongly related to information theory and coding.
- Roughly speaking, the concentration of measure phenomenon can be stated in the following simple way: "A random variable that depends in a smooth way on many independent random variables (but not too much on any of them) is essentially constant" (Talagrand, 1996).

Concentration Inequalities

- Concentration inequalities provide upper bounds on tail probabilities of the type $\mathbb{P}(|X - \overline{x}| \geq t)$ (or $\mathbb{P}(X - \overline{x} \geq t)$ for a random variable (RV) X, where $\overline{x}$ denotes the expectation or median of X.
- Several techniques were developed to prove concentration.

Survey Papers on Concentration Inequalities for Martingales

1. C. McDiarmid, "Concentration," *Probabilistic Methods for Algorithmic Discrete Mathematics*, pp. 195–248, Springer, 1998.
2. F. Chung and L. Lu, "Concentration inequalities and martingale inequalities: a survey," *Internet Mathematics*, vol. 3, no. 1, pp. 79–127, March 2006. Available at http://www.ucsd.edu/∼fan/wp/concen.pdf.
3. I. S., "On Refined Versions of the Azuma-Hoeffding Inequality with Applications in Information Theory, Communications and Coding," a tutorial paper with some original results, July 2011. Last updated in July 2012. See http://arxiv.org/abs/1111.1977.

Definition - [Discrete-Time Martingales]

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots$ be a sequence of sub σ-algebras of $\mathcal{F}$ (which is called a filtration). A sequence $X_0, X_1, \ldots$ of RVs is a martingale if for every i

1. $X_i : \Omega \to \mathbb{R}$ is $\mathcal{F}_i$-measurable, i.e.,

$$\{\omega \in \Omega : X_i(\omega) \leq t\} \in \mathcal{F}_i \quad \forall i \in \{0, 1, \ldots\}, t \in \mathbb{R}.$$

2. $\mathbb{E}[|X_i|] < \infty$.
3. $X_i = \mathbb{E}[X_{i+1}|\mathcal{F}_i]$ almost surely.

Definition - [Discrete-Time Martingales]

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots$ be a sequence of sub σ-algebras of $\mathcal{F}$ (which is called a filtration). A sequence $X_0, X_1, \ldots$ of RVs is a martingale if for every i

1. $X_i : \Omega \to \mathbb{R}$ is $\mathcal{F}_i$-measurable, i.e.,

$$\{\omega \in \Omega : X_i(\omega) \leq t\} \in \mathcal{F}_i \quad \forall i \in \{0, 1, \ldots\}, t \in \mathbb{R}.$$

2. $\mathbb{E}[|X_i|] < \infty$.
3. $X_i = \mathbb{E}[X_{i+1}|\mathcal{F}_i]$ almost surely.

A Simple Example: Random walk

$X_n = \sum_{i=0}^{n} U_i$ where $\mathbb{P}(U_i = +1) = \mathbb{P}(U_i = -1) = \frac{1}{2}$ are i.i.d. RVs.

Martingales – Remarks

Remark 1

Given a RV $X \in \mathbb{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration of sub σ-algebras $\{\mathcal{F}_i\}$, let

$$X_i = \mathbb{E}[X|\mathcal{F}_i] \quad i = 0, 1, \ldots.$$

Then, the sequence $X_0, X_1, \ldots$ forms a martingale.

Martingales – Remarks

Remark 1

Given a RV $X \in \mathbb{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration of sub σ-algebras $\{\mathcal{F}_i\}$, let

$$X_i = \mathbb{E}[X|\mathcal{F}_i] \quad i = 0, 1, \ldots.$$

Then, the sequence $X_0, X_1, \ldots$ forms a martingale.

Remark 2

Choose $\mathcal{F}_0 = \{\Omega, \emptyset\}$ and $\mathcal{F}_n = \mathcal{F}$, then Remark 1 gives a martingale with

$$X_0 = \mathbb{E}[X|\mathcal{F}_0] = \mathbb{E}[X] \quad (\mathcal{F}_0 \text{ doesn't provide information about } X).$$
$$X_n = \mathbb{E}[X|\mathcal{F}_n] = X \text{ a.s.} \quad (\mathcal{F}_n \text{ provides full information about } X).$$

Concentration via the Martingale Approach

Theorem - **[Azuma-Hoeffding inequality]**

Let $\{X_k, \mathcal{F}_k\}_{k=0}^{\infty}$ be a discrete-parameter real-valued martingale. If the jumps of the sequence $\{X_k\}$ are bounded almost surely (a.s.), i.e.,

$$|X_i - X_{i-1}| \leq d_i \quad \forall\, i = 1, 2, \ldots, n \quad \text{a.s.}$$

then

$$\mathbb{P}(|X_n - X_0| \geq r) \leq 2 \exp\left(-\frac{r^2}{2\sum_{i=1}^{n} d_i^2}\right), \quad \forall\, r > 0.$$

Concentration via the Martingale Approach

Theorem - **[Azuma-Hoeffding inequality]**

Let $\{X_k, \mathcal{F}_k\}_{k=0}^{\infty}$ be a discrete-parameter real-valued martingale. If the jumps of the sequence $\{X_k\}$ are bounded almost surely (a.s.), i.e.,

$$|X_i - X_{i-1}| \leq d_i \quad \forall\, i = 1, 2, \ldots, n \quad \text{a.s.}$$

then

$$\mathbb{P}(|X_n - X_0| \geq r) \leq 2 \exp\left(-\frac{r^2}{2\sum_{i=1}^{n} d_i^2}\right), \quad \forall\, r > 0.$$

Concentration Inequalities Around the Average

The Azuma-Hoeffding inequality and Remark 2 enable to get a concentration inequality for X around its expected value ($\mathbb{E}[X]$).

But, the Azuma-Hoeffding inequality is not tight !.
For example, if $r > \sum_{i=1}^{n} d_i \quad \Rightarrow \quad \mathbb{P}(|X_n - X_0| \geq r) = 0.$

Theorem (Th. 2 – McDiarmid '89)

Let $\{X_k, \mathcal{F}_k\}_{k=0}^{\infty}$ be a discrete-parameter real-valued martingale. Assume that, for some constants $d, \sigma > 0$, the following two requirements hold a.s.

$$|X_k - X_{k-1}| \leq d,$$

$$\mathrm{Var}(X_k|\mathcal{F}_{k-1}) = \mathbb{E}\Big[(X_k - X_{k-1})^2 \,\Big|\, \mathcal{F}_{k-1}\Big] \leq \sigma^2$$

for every $k \in \{1, \ldots, n\}$. Then, for every $\alpha \geq 0$,

$$\mathbb{P}(|X_n - X_0| \geq \alpha n) \leq 2 \exp\left(-n\, D\left(\frac{\delta+\gamma}{1+\gamma} \Big\| \frac{\gamma}{1+\gamma}\right)\right)$$

where $\gamma \triangleq \frac{\sigma^2}{d^2}$, $\delta \triangleq \frac{\alpha}{d}$, and $D(p||q) \triangleq p \ln\left(\frac{p}{q}\right) + (1-p)\ln\left(\frac{1-p}{1-q}\right)$.

Corollary

Under the conditions of Theorem 2, for every $\alpha \geq 0$,

$$\mathbb{P}(|X_n - X_0| \geq \alpha n) \leq 2 \exp\left(-n f(\delta)\right)$$

where

$$f(\delta) = \begin{cases} \ln(2)\left[1 - h_2\left(\frac{1-\delta}{2}\right)\right], & 0 \leq \delta \leq 1 \\ +\infty, & \delta > 1 \end{cases}$$

and $h_2(x) \triangleq -x \log_2(x) - (1-x)\log_2(1-x)$ for $0 \leq x \leq 1$.

Corollary

Under the conditions of Theorem 2, for every $\alpha \geq 0$,

$$\mathbb{P}(|X_n - X_0| \geq \alpha n) \leq 2 \exp\left(-n f(\delta)\right)$$

where

$$f(\delta) = \begin{cases} \ln(2)\left[1 - h_2\left(\frac{1-\delta}{2}\right)\right], & 0 \leq \delta \leq 1 \\ +\infty, & \delta > 1 \end{cases}$$

and $h_2(x) \triangleq -x \log_2(x) - (1-x)\log_2(1-x)$ for $0 \leq x \leq 1$.

Proof

Set $\gamma = 1$ in Theorem 2.

Corollary

Under the conditions of Theorem 2, for every $\alpha \geq 0$,

$$\mathbb{P}(|X_n - X_0| \geq \alpha n) \leq 2 \exp\left(-n f(\delta)\right)$$

where

$$f(\delta) = \begin{cases} \ln(2)\left[1 - h_2\left(\frac{1-\delta}{2}\right)\right], & 0 \leq \delta \leq 1 \\ +\infty, & \delta > 1 \end{cases}$$

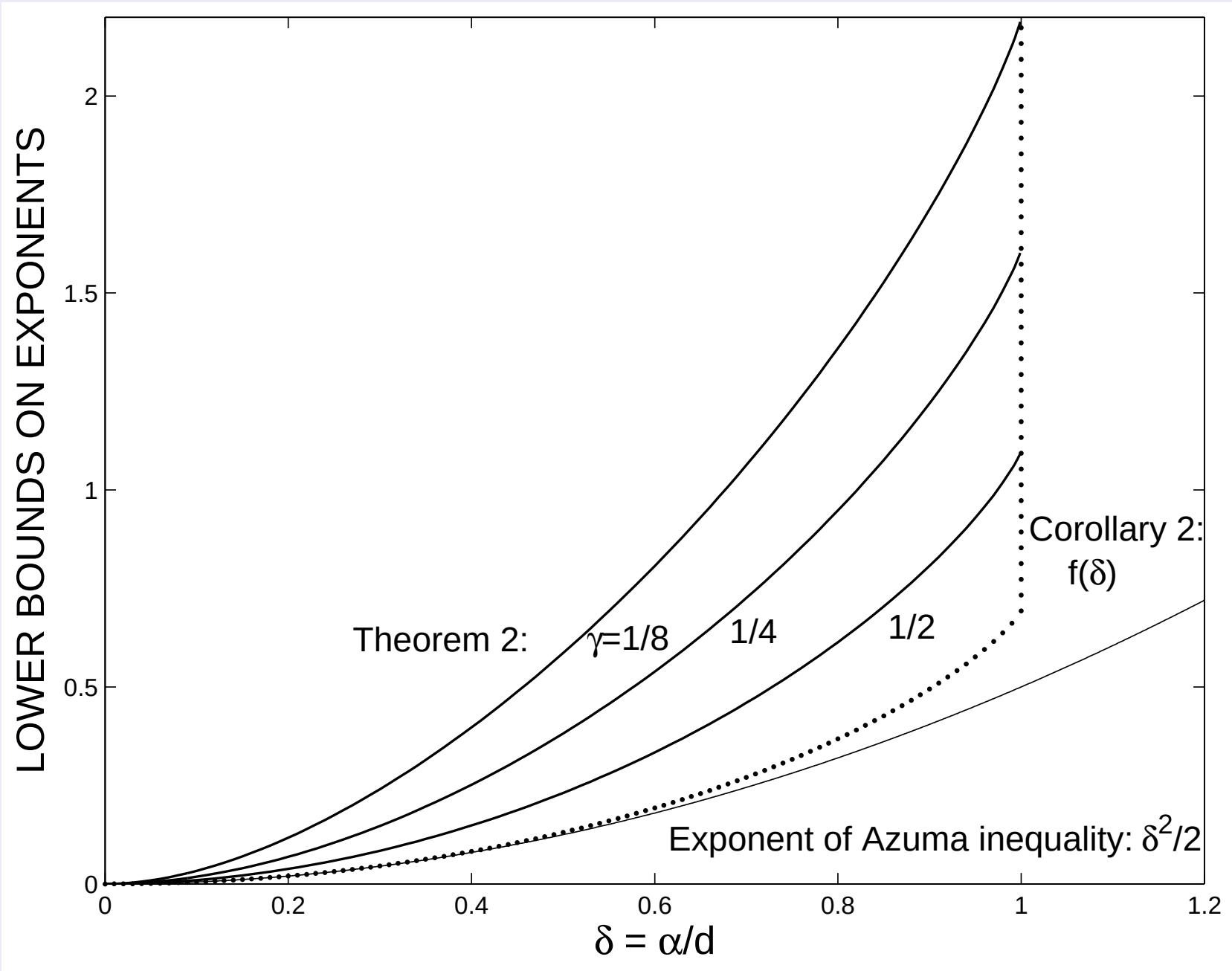
and $h_2(x) \triangleq -x \log_2(x) - (1-x)\log_2(1-x)$ for $0 \leq x \leq 1$.

Proof

Set $\gamma = 1$ in Theorem 2.

Observation (first set $\gamma = 1$, and then use Pinsker's Inequality)

Theorem 2 $\Rightarrow$ Corollary $\Rightarrow$ Azuma-Hoeffding inequality.


LOWER BOUNDS ON EXPONENTS
0
0.5
1
1.5
2
0
0.2
0.4
0.6
0.8
1
1.2
δ = α/d
Theorem 2:
γ=1/8
1/4
1/2
Corollary 2:
f(δ)
Exponent of Azuma inequality: δ²/2

A Simple Example to Motivate Further Improvements

Let

- $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.
- $d > 0$ and $\gamma \in (0, 1]$ be fixed numbers.
- $\{U_k\}_{k \in \mathbb{N}}$ be i.i.d. random variables where

$$\mathbb{P}(U_k = +d) = \mathbb{P}(U_k = -d) = \frac{\gamma}{2}, \quad \mathbb{P}(U_k = 0) = 1 - \gamma, \quad \forall\, k \in \mathbb{N}.$$

A Simple Example to Motivate Further Improvements

Let

- $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.
- $d > 0$ and $\gamma \in (0, 1]$ be fixed numbers.
- $\{U_k\}_{k \in \mathbb{N}}$ be i.i.d. random variables where

$$\mathbb{P}(U_k = +d) = \mathbb{P}(U_k = -d) = \frac{\gamma}{2}, \quad \mathbb{P}(U_k = 0) = 1 - \gamma, \quad \forall\, k \in \mathbb{N}.$$

Consider the Markov process $\{X_k\}_{k \geq 0}$ where $X_0 = 0$ and

$$X_k = X_{k-1} + U_k, \quad \forall k \in \mathbb{N}.$$

A Simple Example (Cont.) - Large Deviation Analysis

From Cramér's theorem in $\mathbb{R}$, for every $\alpha \geq \mathbb{E}[U_1] = 0$,

$$\lim_{n\to\infty} -\frac{1}{n} \ln \mathbb{P}(X_n \geq \alpha n) = I(\alpha)$$

where the rate function is given by

$$I(\alpha) = \sup_{t\geq 0} \{t\alpha - \ln \mathbb{E}[\exp(tU_1)]\}.$$

A Simple Example (Cont.) - Large Deviation Analysis

From Cramér's theorem in $\mathbb{R}$, for every $\alpha \geq \mathbb{E}[U_1] = 0$,

$$\lim_{n\to\infty} -\frac{1}{n} \ln \mathbb{P}(X_n \geq \alpha n) = I(\alpha)$$

where the rate function is given by

$$I(\alpha) = \sup_{t\geq 0} \{t\alpha - \ln \mathbb{E}[\exp(tU_1)]\}.$$

Let $\delta \triangleq \frac{\alpha}{d}$. Calculation shows that

$$I(\alpha) = \delta x - \ln\Big(1 + \gamma\big[\cosh(x) - 1\big]\Big)$$

$$x \triangleq \ln\left(\frac{\delta(1-\gamma) + \sqrt{\delta^2(1-\gamma)^2 + \gamma^2(1-\delta^2)}}{\gamma(1-\delta)}\right).$$

A Simple Example (Cont.) - Large Deviation Analysis

Lets examine the martingale approach in this simple case, where

$$X_k = \sum_{i=1}^{k} U_i, \quad \forall\, k \in \mathbb{N}, \quad X_0 = 0$$

with the natural filtration

$$\mathcal{F}_k = \sigma(U_1, \ldots, U_k), \quad \forall\, k \in \mathbb{N}, \quad \mathcal{F}_0 = \{\emptyset, \Omega\}.$$

A Simple Example (Cont.) - Large Deviation Analysis

Lets examine the martingale approach in this simple case, where

$$X_k = \sum_{i=1}^{k} U_i, \quad \forall\, k \in \mathbb{N}, \quad X_0 = 0$$

with the natural filtration

$$\mathcal{F}_k = \sigma(U_1, \ldots, U_k), \quad \forall\, k \in \mathbb{N}, \quad \mathcal{F}_0 = \{\emptyset, \Omega\}.$$

In this case, the martingale has bounded increments, and for every $k \in \mathbb{N}$

$$|X_k - X_{k-1}| \leq d$$
$$\mathbb{E}[(X_k - X_{k-1})^2 \,|\, \mathcal{F}_{k-1}] = \mathsf{Var}(U_k) = \gamma d^2$$

almost surely.

A Simple Example (Cont.) - Large Deviation Analysis

Via the martingale approach, the following exponential inequality follows:

$$\mathbb{P}\left(X_n - X_0 \geq \alpha n\right) \leq \exp\left(-n\, D\left(\frac{\delta+\gamma}{1+\gamma}\Big|\Big|\frac{\gamma}{1+\gamma}\right)\right)$$

where $\delta \triangleq \frac{\alpha}{d}$, and

$$D(p||q) \triangleq p\ln\left(\frac{p}{q}\right) + (1-p)\ln\left(\frac{1-p}{1-q}\right), \quad \forall\, p, q \in [0,1]$$

is the divergence (relative entropy) between the probability distributions $(p, 1-p)$ and $(q, 1-q)$. If $\delta > 1$, then the above probability is zero.

A Simple Example (Cont.) - Large Deviation Analysis

Via the martingale approach, the following exponential inequality follows:

$$\mathbb{P}\left(X_n - X_0 \geq \alpha n\right) \leq \exp\left(-n\, D\left(\frac{\delta+\gamma}{1+\gamma}\Big|\Big|\frac{\gamma}{1+\gamma}\right)\right)$$

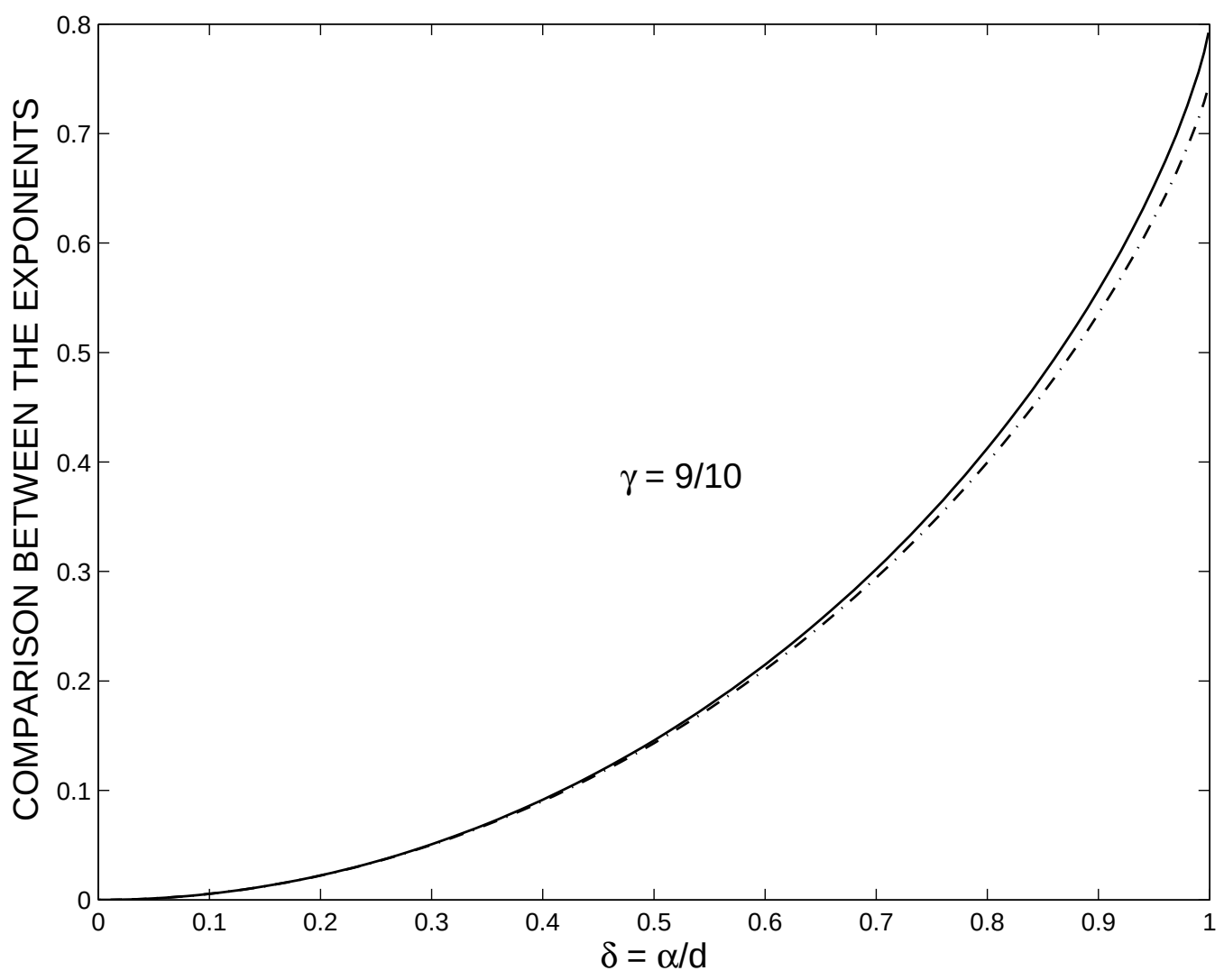
where $\delta \triangleq \frac{\alpha}{d}$, and

$$D(p||q) \triangleq p\ln\left(\frac{p}{q}\right) + (1-p)\ln\left(\frac{1-p}{1-q}\right), \quad \forall\, p, q \in [0,1]$$

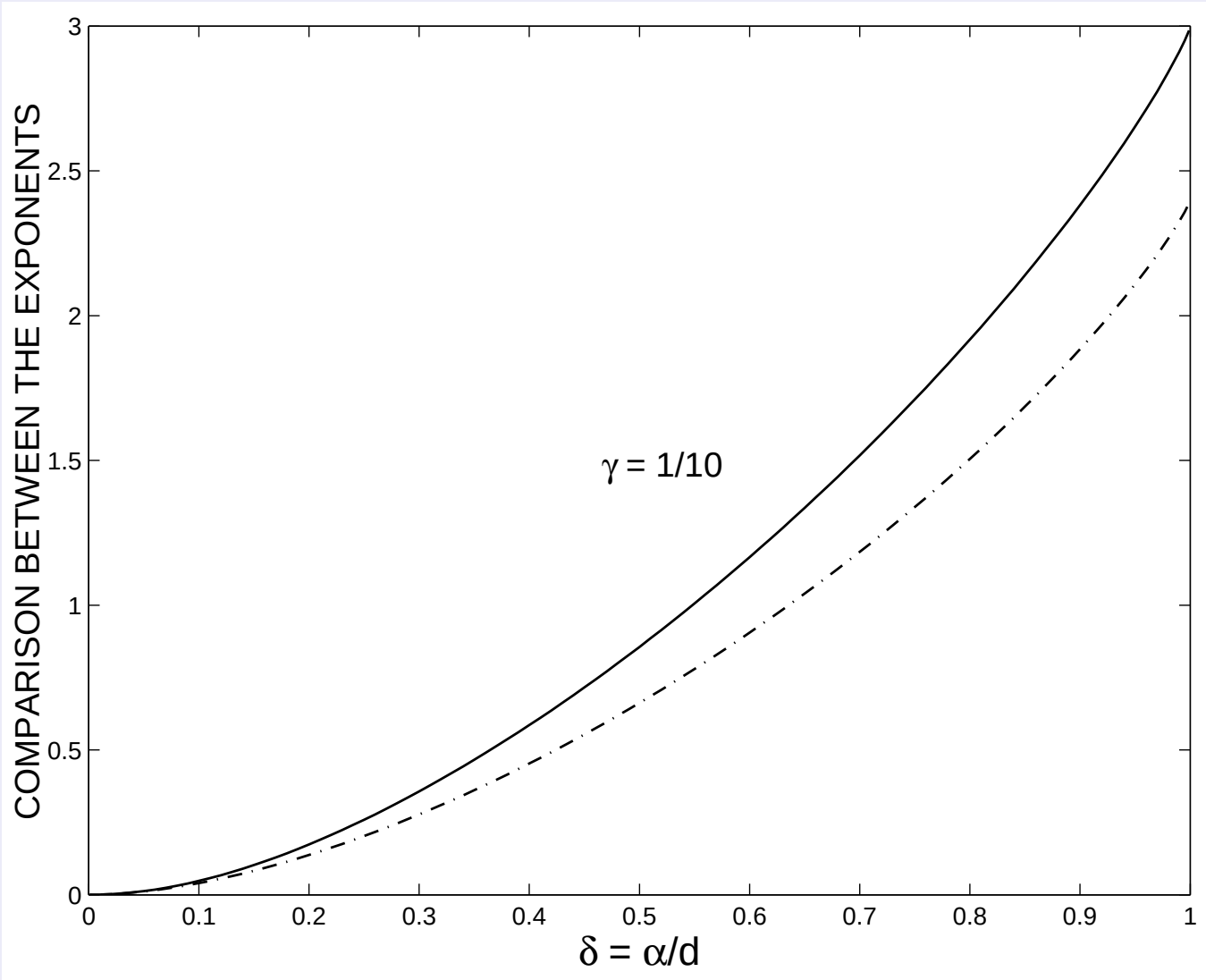
is the divergence (relative entropy) between the probability distributions $(p, 1-p)$ and $(q, 1-q)$. If $\delta > 1$, then the above probability is zero.

This exponential bound is not tight (unless $\gamma = 1$), and the gap to the exact asymptotic exponent increases as the value of $\gamma \in (0,1]$ is decreased.

Comparison of the Exact Exponent and Its Lower Bound



Comparison of the Exact Exponent and Its Lower Bound (Cont.)



Questions:

- Why the lower bound on the exponent is not asymptotically tight ?
- Why this gap increases as the value of γ is decreased ?
- How this situation can be improved via the martingale approach ?

Questions:

- Why the lower bound on the exponent is not asymptotically tight ?
- Why this gap increases as the value of γ is decreased ?
- How this situation can be improved via the martingale approach ?

This exponential bound relies on Bennett's inequality: Since $U_k \leq d$, $\mathbb{E}[U_k] = 0$, and $\mathrm{Var}(U_k) = \gamma d^2$, then

$$\mathbb{E}\left[\exp(tU_k)\right] \leq \frac{\gamma \exp(td) + \exp(-\gamma td)}{1+\gamma}, \quad \forall\, t \geq 0.$$

Questions:

- Why the lower bound on the exponent is not asymptotically tight ?
- Why this gap increases as the value of γ is decreased ?
- How this situation can be improved via the martingale approach ?

This exponential bound relies on Bennett's inequality: Since $U_k \leq d$, $\mathbb{E}[U_k] = 0$, and $\mathrm{Var}(U_k) = \gamma d^2$, then

$$\mathbb{E}\left[\exp(tU_k)\right] \leq \frac{\gamma \exp(td) + \exp(-\gamma td)}{1+\gamma}, \quad \forall\, t \geq 0.$$

The probability distribution that achieves Bennett's inequality with equality is asymmetric (unless $\gamma = 1$), and is equal to

$$\mathbb{P}(\tilde{U}_k = d) = \frac{\gamma}{1+\gamma}, \quad \mathbb{P}(\tilde{U}_k = -\gamma d) = \frac{1}{1+\gamma}.$$

By reducing the value of $\gamma \in (0, 1]$, the above asymmetry grows. Indeed, this enlarges the gap to the exact asymptotic exponent.

Interim Conclusion

An improvement is likely to be obtained by a refinement of Bennett's inequality when X_k is conditionally symmetric given $\mathcal{F}_{k-1}$.

Interim Conclusion

An improvement is likely to be obtained by a refinement of Bennett's inequality when X_k is conditionally symmetric given $\mathcal{F}_{k-1}$.

Definition: Conditionally Symmetric Martingales

Let $\{X_k, \mathcal{F}_k\}_{k\in\mathbb{N}_0}$, where $\mathbb{N}_0 \triangleq \mathbb{N} \cup \{0\}$, be a discrete-time and real-valued martingale, and let $\xi_k \triangleq X_k - X_{k-1}$ for every $k \in \mathbb{N}$. Then $\{X_k, \mathcal{F}_k\}_{k\in\mathbb{N}_0}$ is a *conditionally symmetric martingale* if, conditioned on $\mathcal{F}_{k-1}$, the RV ξ_k is symmetrically distributed around zero.

Interim Conclusion

An improvement is likely to be obtained by a refinement of Bennett's inequality when X_k is conditionally symmetric given $\mathcal{F}_{k-1}$.

Definition: Conditionally Symmetric Martingales

Let $\{X_k, \mathcal{F}_k\}_{k\in\mathbb{N}_0}$, where $\mathbb{N}_0 \triangleq \mathbb{N}\cup\{0\}$, be a discrete-time and real-valued martingale, and let $\xi_k \triangleq X_k - X_{k-1}$ for every $k \in \mathbb{N}$. Then $\{X_k, \mathcal{F}_k\}_{k\in\mathbb{N}_0}$ is a *conditionally symmetric martingale* if, conditioned on $\mathcal{F}_{k-1}$, the RV ξ_k is symmetrically distributed around zero.

Definition: Conditionally Symmetric Sub/ Supermartingales

Let $\{X_k, \mathcal{F}_k\}_{k\in\mathbb{N}_0}$ be a discrete-time real-valued sub or supermartingale, and let $\eta_k \triangleq X_k - \mathbb{E}[X_k|\mathcal{F}_{k-1}]$ for every $k \in \mathbb{N}$. Then it is conditionally symmetric if, conditioned on $\mathcal{F}_{k-1}$, the RV η_k is symmetrically distributed around zero.

Construction of Conditionally Symmetric Martingales

Example 1: Let

- $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space
- $\{U_k\}_{k\in\mathbb{N}} \subseteq L^1(\Omega, \mathcal{F}, \mathbb{P})$ be independent random variables that are symmetrically distributed around zero
- $\{\mathcal{F}_k\}_{k\geq 0}$ be the natural filtration where $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_k = \sigma(U_1, \ldots, U_k),\ \forall\, k \in \mathbb{N}$.
- For $k \in \mathbb{N}$, let $A_k \in L^\infty(\Omega, \mathcal{F}_{k-1}, \mathbb{P})$ be an $\mathcal{F}_{k-1}$-measurable random variable with a finite essential supremum.

Construction of Conditionally Symmetric Martingales

Example 1: Let

- $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space
- $\{U_k\}_{k\in\mathbb{N}} \subseteq L^1(\Omega, \mathcal{F}, \mathbb{P})$ be independent random variables that are symmetrically distributed around zero
- $\{\mathcal{F}_k\}_{k\geq 0}$ be the natural filtration where $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_k = \sigma(U_1, \ldots, U_k),\ \forall\, k \in \mathbb{N}$.
- For $k \in \mathbb{N}$, let $A_k \in L^\infty(\Omega, \mathcal{F}_{k-1}, \mathbb{P})$ be an $\mathcal{F}_{k-1}$-measurable random variable with a finite essential supremum.

Define a new sequence of random variables in $L^1(\Omega, \mathcal{F}, \mathbb{P})$ where

$$X_n = \sum_{k=1}^{n} A_k U_k, \quad \forall\, n \in \mathbb{N}$$

and $X_0 = 0$. Then, $\{X_n, \mathcal{F}_n\}_{n\in\mathbb{N}_0}$ is a conditionally symmetric martingale.

Construction of Conditionally Symmetric Martingales

Example 2: Let

- $\{X_n, \mathcal{F}_n\}_{n\in\mathbb{N}_0}$ be a conditionally symmetric martingale
- $A_k \in L^\infty(\Omega, \mathcal{F}_{k-1}, \mathbb{P})$ be an $\mathcal{F}_{k-1}$-measurable random variable with a finite essential supremum.

Define $Y_0 = 0$ and $Y_n = \sum_{k=1}^n A_k(X_k - X_{k-1}), \quad \forall\, n \in \mathbb{N}$.
Then, $\{Y_n, \mathcal{F}_n\}_{n\in\mathbb{N}_0}$ is a conditionally symmetric martingale.

Construction of Conditionally Symmetric Martingales

Example 2: Let

- $\{X_n, \mathcal{F}_n\}_{n\in\mathbb{N}_0}$ be a conditionally symmetric martingale
- $A_k \in L^\infty(\Omega, \mathcal{F}_{k-1}, \mathbb{P})$ be an $\mathcal{F}_{k-1}$-measurable random variable with a finite essential supremum.

Define $Y_0 = 0$ and $Y_n = \sum_{k=1}^n A_k(X_k - X_{k-1}), \quad \forall\, n \in \mathbb{N}$.
Then, $\{Y_n, \mathcal{F}_n\}_{n\in\mathbb{N}_0}$ is a conditionally symmetric martingale.

Example 3: A sampled Brownian motion is a discrete time, conditionally symmetric martingale with un-bounded increments.

Construction of Conditionally Symmetric Martingales

Example 2: Let

- $\{X_n, \mathcal{F}_n\}_{n\in\mathbb{N}_0}$ be a conditionally symmetric martingale
- $A_k \in L^\infty(\Omega, \mathcal{F}_{k-1}, \mathbb{P})$ be an $\mathcal{F}_{k-1}$-measurable random variable with a finite essential supremum.

Define $Y_0 = 0$ and $Y_n = \sum_{k=1}^n A_k(X_k - X_{k-1}), \quad \forall\, n \in \mathbb{N}$.
Then, $\{Y_n, \mathcal{F}_n\}_{n\in\mathbb{N}_0}$ is a conditionally symmetric martingale.

Example 3: A sampled Brownian motion is a discrete time, conditionally symmetric martingale with un-bounded increments.

Goal: Our next goal is to demonstrate how the assumption of the conditional symmetry improves existing exponential inequalities for discrete-time real-valued martingales with bounded increments.

Exponential Inequalities for Conditionally Symmetric Martingales

Lemma

Let X be a real-valued RV with a symmetric distribution around zero, a support $[-d, d]$, and assume Var$(X) \leq \gamma d^2$ for some $d > 0$ and $\gamma \in [0, 1]$. Let h be a real-valued convex function, and assume that $h(d^2) \geq h(0)$. Then,

$$\mathbb{E}[h(X^2)] \leq (1-\gamma)h(0) + \gamma h(d^2)$$

with equality if $\mathbb{P}(X = \pm d) = \frac{\gamma}{2}, \quad \mathbb{P}(X = 0) = 1 - \gamma.$

Exponential Inequalities for Conditionally Symmetric Martingales

Lemma

Let X be a real-valued RV with a symmetric distribution around zero, a support $[-d, d]$, and assume $Var(X) \leq \gamma d^2$ for some $d > 0$ and $\gamma \in [0, 1]$. Let h be a real-valued convex function, and assume that $h(d^2) \geq h(0)$. Then,

$$\mathbb{E}[h(X^2)] \leq (1-\gamma)h(0) + \gamma h(d^2)$$

with equality if $\mathbb{P}(X = \pm d) = \frac{\gamma}{2}, \quad \mathbb{P}(X = 0) = 1 - \gamma.$

Proof

- h convex, $X \in [-d, d]$ a.s. $\Rightarrow h(X^2) \leq h(0) + \left(\frac{X}{d}\right)^2 \left(h(d^2) - h(0)\right)$.
- Taking expectations on both sides gives the inequality, with an equality for the above symmetric distribution.

Exp. Inequalities for Conditionally Symmetric Martingales (Cont.)

Corollary

Let X be a real-valued RV with a symmetric distribution around zero, a support $[-d, d]$, and assume $\mathrm{Var}(X) \leq \gamma d^2$ for some $d > 0$ and $\gamma \in [0, 1]$. Then,

$$\mathbb{E}\big[\exp(\lambda X)\big] \leq 1 + \gamma\big[\cosh(\lambda d) - 1\big], \quad \forall\, \lambda \in \mathbb{R}$$

with equality if $\mathbb{P}(X = \pm d) = \frac{\gamma}{2}, \quad \mathbb{P}(X = 0) = 1 - \gamma$.

Exp. Inequalities for Conditionally Symmetric Martingales (Cont.)

Corollary

Let X be a real-valued RV with a symmetric distribution around zero, a support $[-d, d]$, and assume $\mathrm{Var}(X) \leq \gamma d^2$ for some $d > 0$ and $\gamma \in [0, 1]$. Then,

$$\mathbb{E}\big[\exp(\lambda X)\big] \leq 1 + \gamma\big[\cosh(\lambda d) - 1\big], \quad \forall\, \lambda \in \mathbb{R}$$

with equality if $\mathbb{P}(X = \pm d) = \frac{\gamma}{2}, \quad \mathbb{P}(X = 0) = 1 - \gamma.$

Proof

- Symmetric distribution of $X \quad \Rightarrow \quad \mathbb{E}\big[\exp(\lambda X)\big] = \mathbb{E}\big[\cosh(\lambda X)\big].$
- The corollary follows from the lemma since, for every $x \in \mathbb{R}$, $\cosh(\lambda x) = h(x^2)$ where $h(x) \triangleq \sum_{n=0}^{\infty} \frac{\lambda^{2n} |x|^n}{(2n)!}$ is a convex function, and $h(d^2) = \cosh(\lambda d) \geq 1 = h(0)$.

Theorem 2 (I. S., 2012)

Let $\{X_k, \mathcal{F}_k\}_{k\in\mathbb{N}_0}$ be a discrete-time real-valued and conditionally symmetric martingale. Assume that, for some fixed numbers $d, \sigma > 0$, the following two requirements are satisfied a.s.

$$|X_k - X_{k-1}| \le d, \qquad \mathsf{Var}(X_k|\mathcal{F}_{k-1}) = \mathbb{E}\big[(X_k - X_{k-1})^2 \,|\, \mathcal{F}_{k-1}\big] \le \sigma^2$$

for every $k \in \mathbb{N}$.

Theorem 2 (I. S., 2012)

Let $\{X_k, \mathcal{F}_k\}_{k\in\mathbb{N}_0}$ be a discrete-time real-valued and conditionally symmetric martingale. Assume that, for some fixed numbers $d, \sigma > 0$, the following two requirements are satisfied a.s.

$$|X_k - X_{k-1}| \leq d, \qquad \mathrm{Var}(X_k|\mathcal{F}_{k-1}) = \mathbb{E}\big[(X_k - X_{k-1})^2 \,|\, \mathcal{F}_{k-1}\big] \leq \sigma^2$$

for every $k \in \mathbb{N}$. Then, for every $\alpha \geq 0$ and $n \in \mathbb{N}$,

$$\mathbb{P}\left(\max_{1\leq k\leq n} |X_k - X_0| \geq \alpha n\right) \leq 2\exp\big(-nE(\gamma,\delta)\big)$$

where

$$\gamma \triangleq \frac{\sigma^2}{d^2}, \quad \delta \triangleq \frac{\alpha}{d}$$

and for $\gamma \in (0, 1]$ and $\delta \in [0, 1)$, the exponent $E(\gamma, \delta)$ is given as follows:

Theorem 2 (Cont.)

$$E(\gamma,\delta) \triangleq \delta x - \ln\Big(1+\gamma\big[\cosh(x)-1\big]\Big)$$

$$x \triangleq \ln\left(\frac{\delta(1-\gamma)+\sqrt{\delta^2(1-\gamma)^2+\gamma^2(1-\delta^2)}}{\gamma(1-\delta)}\right).$$

If $\delta > 1$, then the probability is zero (so $E(\gamma,\delta) \triangleq +\infty$), and $E(\gamma,1) = \ln\big(\frac{2}{\gamma}\big)$.

Theorem 2 (Cont.)

$$E(\gamma,\delta) \triangleq \delta x - \ln\Big(1+\gamma\big[\cosh(x)-1\big]\Big)$$

$$x \triangleq \ln\left(\frac{\delta(1-\gamma)+\sqrt{\delta^2(1-\gamma)^2+\gamma^2(1-\delta^2)}}{\gamma(1-\delta)}\right).$$

If $\delta > 1$, then the probability is zero (so $E(\gamma,\delta) \triangleq +\infty$), and $E(\gamma,1) = \ln\big(\frac{2}{\gamma}\big)$.

Asymptotic Optimality of the Exponent

The example we studied earlier shows that the exponent of the bound in Theorem 1 is asymptotically optimal. That is, there exists a conditionally symmetric martingale, satisfying the conditions in this theorem, that attains the exponent $E(\gamma,\delta)$ in the limit where $n \to \infty$.

Theorem 2 should be compared to the bound in Theorem 1 which does not require the conditional symmetry property.

Theorem 1 (Reminder)

[McDiarmid 1989, book of Dembo & Zeitouni (Corollary 2.4.7)]

Let $\{X_k, \mathcal{F}_k\}_{k \in \mathbb{N}_0}$ be a discrete-time real-valued martingale with bounded jumps. Assume that the two conditions on the bounded increments and conditional variance from Theorem 1 are satisfied a.s. for every $k \in \mathbb{N}$.

Theorem 2 should be compared to the bound in Theorem 1 which does not require the conditional symmetry property.

Theorem 1 (Reminder)

[McDiarmid 1989, book of Dembo & Zeitouni (Corollary 2.4.7)]

Let $\{X_k, \mathcal{F}_k\}_{k\in\mathbb{N}_0}$ be a discrete-time real-valued martingale with bounded jumps. Assume that the two conditions on the bounded increments and conditional variance from Theorem 1 are satisfied a.s. for every $k \in \mathbb{N}$. Then, for every $\alpha \geq 0$ and $n \in \mathbb{N}$,

$$\mathbb{P}\left(\max_{1\leq k\leq n} |X_k - X_0| \geq \alpha n\right) \leq 2\exp\left(-n\, D\left(\frac{\delta+\gamma}{1+\gamma}\Big|\Big|\frac{\gamma}{1+\gamma}\right)\right)$$

where $\gamma \triangleq \frac{\sigma^2}{d^2}$ and $\delta \triangleq \frac{\alpha}{d}$ were introduced earlier, and

$$D(p||q) \triangleq p\ln\left(\frac{p}{q}\right) + (1-p)\ln\left(\frac{1-p}{1-q}\right), \quad \forall\, p, q \in [0,1].$$

If $\delta > 1$, then the probability is zero.

Proof Technique for Theorems 1 and 2

- Define $X_n - X_0 = \sum_{k=1}^{n} \xi_k$ where $\xi_k \triangleq X_k - X_{k-1}$.

Proof Technique for Theorems 1 and 2

- Define $X_n - X_0 = \sum_{k=1}^{n} \xi_k$ where $\xi_k \triangleq X_k - X_{k-1}$.
- $|\xi_k| \leq d$, $\mathbb{E}\left[\xi_k \,|\, \mathcal{F}_{k-1}\right] = 0$, $\mathsf{Var}(\xi_k \,|\, \mathcal{F}_{k-1}) \leq \gamma d^2$. The RV ξ_k is $\mathcal{F}_k$-measurable.

Proof Technique for Theorems 1 and 2

- Define $X_n - X_0 = \sum_{k=1}^{n} \xi_k$ where $\xi_k \triangleq X_k - X_{k-1}$.
- $|\xi_k| \leq d$, $\mathbb{E}\big[\xi_k \,|\, \mathcal{F}_{k-1}\big] = 0$, $\mathsf{Var}(\xi_k \,|\, \mathcal{F}_{k-1}) \leq \gamma d^2$. The RV ξ_k is $\mathcal{F}_k$-measurable.
- Exponentiation and the maximal inequality for submartingales give that, for every $\alpha \geq 0$,

$$\mathbb{P}\left(\max_{1\leq k\leq n}(X_k - X_0) \geq n\alpha\right) \leq e^{-n\alpha t}\,\mathbb{E}\left[\exp\left(t\sum_{k=1}^{n}\xi_k\right)\right], \quad \forall\, t \geq 0.$$

Proof Technique for Theorems 1 and 2

- Define $X_n - X_0 = \sum_{k=1}^{n} \xi_k$ where $\xi_k \triangleq X_k - X_{k-1}$.
- $|\xi_k| \leq d$, $\mathbb{E}\big[\xi_k \,|\, \mathcal{F}_{k-1}\big] = 0$, $\mathrm{Var}(\xi_k \,|\, \mathcal{F}_{k-1}) \leq \gamma d^2$. The RV ξ_k is $\mathcal{F}_k$-measurable.
- Exponentiation and the maximal inequality for submartingales give that, for every $\alpha \geq 0$,

$$\mathbb{P}\left(\max_{1\leq k\leq n}(X_k - X_0) \geq n\alpha\right) \leq e^{-n\alpha t}\,\mathbb{E}\left[\exp\left(t\sum_{k=1}^{n}\xi_k\right)\right], \quad \forall\, t \geq 0.$$

- Since $\{\mathcal{F}_i\}$ forms a filtration, then for every $t \geq 0$

$$\mathbb{E}\left[\exp\left(t\sum_{k=1}^{n}\xi_k\right)\right] = \mathbb{E}\left[\exp\left(t\sum_{k=1}^{n-1}\xi_k\right)\mathbb{E}\big[\exp(t\xi_n)\,|\,\mathcal{F}_{n-1}\big]\right].$$

Proof Technique for Theorems 1 and 2 (Cont.)

- For proving Theorem 2, the use of Bennett's inequality gives

$$\mathbb{E}\left[\exp(t\xi_k)\,|\,\mathcal{F}_{k-1}\right] \leq \frac{\gamma\exp(td)+\exp(-\gamma td)}{1+\gamma}.$$

Proof Technique for Theorems 1 and 2 (Cont.)

- For proving Theorem 2, the use of Bennett's inequality gives

$$\mathbb{E}\left[\exp(t\xi_k)\,|\,\mathcal{F}_{k-1}\right] \leq \frac{\gamma\exp(td)+\exp(-\gamma td)}{1+\gamma}.$$

- For the refinement in Theorem 1 for conditionally symmetric martingales with bounded increments, use of Lemma 1 gives

$$\mathbb{E}\left[\exp(t\xi_k)\,|\,\mathcal{F}_{k-1}\right] \leq 1+\gamma\Big[\cosh(td)-1\Big].$$

Proof Technique for Theorems 1 and 2 (Cont.)

- For proving Theorem 2, the use of Bennett's inequality gives

$$\mathbb{E}\left[\exp(t\xi_k)\,|\,\mathcal{F}_{k-1}\right] \leq \frac{\gamma\exp(td)+\exp(-\gamma td)}{1+\gamma}.$$

- For the refinement in Theorem 1 for conditionally symmetric martingales with bounded increments, use of Lemma 1 gives

$$\mathbb{E}\left[\exp(t\xi_k)\,|\,\mathcal{F}_{k-1}\right] \leq 1+\gamma\Big[\cosh(td)-1\Big].$$

- In both cases, one gets exponential bounds with the free parameter t.

Proof Technique for Theorems 1 and 2 (Cont.)

- For proving Theorem 2, the use of Bennett's inequality gives

$$\mathbb{E}\left[\exp(t\xi_k)\,|\,\mathcal{F}_{k-1}\right] \leq \frac{\gamma\exp(td)+\exp(-\gamma td)}{1+\gamma}.$$

- For the refinement in Theorem 1 for conditionally symmetric martingales with bounded increments, use of Lemma 1 gives

$$\mathbb{E}\left[\exp(t\xi_k)\,|\,\mathcal{F}_{k-1}\right] \leq 1+\gamma\big[\cosh(td)-1\big].$$

- In both cases, one gets exponential bounds with the free parameter t.
- Optimization over $t \geq 0$ & union bound to get two-sided inequalities.

Proof Technique for Theorems 1 and 2 (Cont.)

- For proving Theorem 2, the use of Bennett's inequality gives

$$\mathbb{E}\left[\exp(t\xi_k)\,|\,\mathcal{F}_{k-1}\right] \leq \frac{\gamma\exp(td)+\exp(-\gamma td)}{1+\gamma}.$$

- For the refinement in Theorem 1 for conditionally symmetric martingales with bounded increments, use of Lemma 1 gives

$$\mathbb{E}\left[\exp(t\xi_k)\,|\,\mathcal{F}_{k-1}\right] \leq 1+\gamma\Big[\cosh(td)-1\Big].$$

- In both cases, one gets exponential bounds with the free parameter t.
- Optimization over $t \geq 0$ & union bound to get two-sided inequalities.
- The asymptotic optimality of the exponent in Theorem 1 follows from the simple example that was shown earlier.

Relation to Classical Results in Probability Theory:

The above concentration inequalities are linked to

- Central limit theorem for martingales
- Law of iterated logarithm
- Moderate deviations principle for real-valued i.i.d. RVs.

For proofs and discussions on these relations, see Section IV in http://arxiv.org/abs/1111.1977.

Binary Hypothesis Testing

Let the RVs $X_1, X_2....$ be i.i.d. $\sim Q$, and consider two hypotheses:

- $H_1 : Q = P_1$.
- $H_2 : Q = P_2$.

For simplicity, assume that the RVs are discrete, and take their values on a finite alphabet $\mathcal{X}$ where $P_1(x), P_2(x) > 0$ for every $x \in \mathcal{X}$.

The log-likelihood ratio (LLR): $L(X_1^n) = \sum_{i=1}^n \ln \frac{P_1(X_i)}{P_2(X_i)}$.

By the strong law of large numbers (SLLN), if H_1 is true, then a.s. $\lim_{n\to\infty} \frac{L(X_1^n)}{n} = D(P_1||P_2)$ and, if H_2, $\lim_{n\to\infty} \frac{L(X_1^n)}{n} = -D(P_2||P_1)$.

Let $\overline{\lambda}, \underline{\lambda} \in \mathbb{R}$ satisfy $-D(P_2||P_1) < \underline{\lambda} \leq \overline{\lambda} < D(P_1||P_2)$.
Decide on H_1 if $L(X_1^n) > n\overline{\lambda}$ and on H_2 if $L(X_1^n) < n\underline{\lambda}$.

Let

$$\alpha_n^{(1)} \triangleq P_1^n\left(L(X_1^n) \leq n\overline{\lambda}\right) \qquad \alpha_n^{(2)} \triangleq P_1^n\left(L(X_1^n) \leq n\underline{\lambda}\right)$$

and

$$\beta_n^{(1)} \triangleq P_2^n\left(L(X_1^n) \geq n\underline{\lambda}\right) \qquad \beta_n^{(2)} \triangleq P_2^n\left(L(X_1^n) \geq n\overline{\lambda}\right)$$

then

- $\alpha_n^{(1)}$ and $\beta_n^{(1)}$ are the probabilities of either making an error/ declaring an erasure under, respectively, hypotheses H_1 and H_2.
- $\alpha_n^{(2)}$ and $\beta_n^{(2)}$ are the probabilities of making an error under, respectively, hypotheses H_1 and H_2.

Large deviations analysis $\Rightarrow$ α_n and β_n both decay exponentially to zero as a function of the block length (n).

Moderate-Deviations Analysis for Binary Hypothesis Testing

- Instead of the setting where the thresholds are kept fixed independently of n, let these thresholds tend to their asymptotic limits (due to the SLLN), i.e.,

$$\lim_{n\to\infty} \overline{\lambda}^{(n)} = D(P_1||P_2), \quad \lim_{n\to\infty} \underline{\lambda}^{(n)} = -D(P_2||P_1).$$

- In moderate-deviations analysis of binary hypothesis testing, we are interested to analyze the case where
  1. The block length n of the input sequence tends to infinity.
  2. The thresholds tend to their asymptotic limits

  simultaneously.

Moderate-Deviations Analysis for Binary Hypothesis Testing (Cont.)

To this end, let $\eta \in (\frac{1}{2}, 1)$, and $\varepsilon_1, \varepsilon_2 > 0$ be arbitrary fixed numbers. Set the upper and lower thresholds to

$$\overline{\lambda}^{(n)} = D(P_1||P_2) - \varepsilon_1 n^{-(1-\eta)}$$
$$\underline{\lambda}^{(n)} = -D(P_2||P_1) + \varepsilon_2 n^{-(1-\eta)}.$$

Moderate-Deviations Analysis via the Martingale Approach

Under hypothesis H_1, construct the martingale $\{U_k, \mathcal{F}_k\}_{k=0}^n$ where

- $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \mathcal{F}_n$ is the natural filtration

$$\mathcal{F}_0 = \{\emptyset, \Omega\}, \quad \mathcal{F}_k = \sigma(X_1, \ldots, X_k), \ \ \forall\, k \in \{1, \ldots, n\}.$$

- $U_k = \mathbb{E}_{P_1^n}\big[L(X_1^n) \mid \mathcal{F}_k\big]$.

Moderate-Deviations Analysis via the Martingale Approach

Under hypothesis H_1, construct the martingale $\{U_k, \mathcal{F}_k\}_{k=0}^{n}$ where

- $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \mathcal{F}_n$ is the natural filtration

$$\mathcal{F}_0 = \{\emptyset, \Omega\}, \quad \mathcal{F}_k = \sigma(X_1, \ldots, X_k), \ \ \forall\, k \in \{1, \ldots, n\}.$$

- $U_k = \mathbb{E}_{P_1^n}\big[L(X_1^n) \mid \mathcal{F}_k\big]$.

For every $k \in \{0, \ldots, n\}$: $\quad U_k = \sum_{i=1}^{k} \ln \frac{P_1(X_i)}{P_2(X_i)} + (n-k) D(P_1 || P_2)$.

Moderate-Deviations Analysis via the Martingale Approach (Cont.)

- $U_0 = nD(P_1||P_2)$, and $U_n = L(X_1^n)$.
- Let

$$d_1 \triangleq \max_{x\in\mathcal{X}} \left| \ln \frac{P_1(x)}{P_2(x)} - D(P_1||P_2) \right| \quad \Rightarrow d_1 < \infty.$$

$\Rightarrow |U_k - U_{k-1}| \leq d_1$ holds a.s. for every $k \in \{1, \ldots, n\}$.

Due to the independence of the RVs in the sequence $\{X_i\}$

$$\begin{aligned}
&\mathbb{E}_{P_1^n}\left[(U_k - U_{k-1})^2 \,|\, \mathcal{F}_{k-1}\right] \\
&= \sum_{x\in\mathcal{X}} \left\{ P_1(x) \left(\ln \frac{P_1(x)}{P_2(x)} - D(P_1||P_2) \right)^2 \right\} \\
&\triangleq \sigma_1^2.
\end{aligned}$$

Moderate-Deviations Analysis via the Martingale Approach (Cont.)

- Let $\varepsilon_1 > 0$ and $\eta \in (\frac{1}{2}, 1)$ be two arbitrarily fixed numbers as above.
- Under hypothesis H_1, it follows from Theorem 2 and the above construction of a martingale that

$$P_1^n\big(L(X_1^n) \leq n\overline{\lambda}^{(n)}\big) \leq \exp\left(-nD\left(\frac{\delta_1^{(\eta,n)} + \gamma_1}{1+\gamma_1} \,\|\, \frac{\gamma_1}{1+\gamma_1}\right)\right)$$

where

$$\delta_1^{(\eta,n)} \triangleq \frac{\varepsilon_1 n^{-(1-\eta)}}{d_1}, \quad \gamma_1 \triangleq \frac{\sigma_1^2}{d_1^2}$$

with d_1 and σ_1^2 as defined in the previous slide.

Moderate-Deviations Analysis via the Martingale Approach (Cont.)

- The concentration inequality in Theorem 2 and a simple inequality give that, under hypothesis H_1, for every $\eta \in \left(\frac{1}{2}, 1\right)$,

$$\alpha_n^{(1)} \leq \exp\left(-\frac{\varepsilon_1^2 \, n^{2\eta-1}}{2\sigma_1^2}\left(1 - \frac{\varepsilon_1 d_1}{3\sigma_1^2(1+\gamma_1)} \, \frac{1}{n^{1-\eta}}\right)\right)$$

  so this upper bound on the overall probability of either making an error or declaring an erasure under hypothesis H_1 decays sub-exponentially to zero. It improves by increasing $\eta \in (\frac{1}{2}, 1)$.
- On the other hand, the exponential decay of the probability of error $\alpha_n^{(2)}$ improves as the value of $\eta \in (\frac{1}{2}, 1)$ is decreased (since the margin for making an error is increased).

Moderate-Deviations Analysis via the Martingale Approach (Cont.)

Moderate-deviations analysis reflects, as can be expected, a tradeoff between the two α_n's and also between the two β_n's. But all decay asymptotically to zero as n tends to infinity (which is indeed consistent with the SLLN).

Moderate-Deviations Analysis via the Martingale Approach (Cont.)

The following upper bound implies that

$$\lim_{n\to\infty} n^{1-2\eta} \ \ln \ P_1^n\Big(L(X_1^n) \leq n\overline{\lambda}^{(n)}\Big) \leq -\frac{\varepsilon_1^2}{2\sigma_1^2}.$$

Moderate-Deviations Analysis via the Martingale Approach (Cont.)

The following upper bound implies that

$$\lim_{n\to\infty} n^{1-2\eta} \ \ln \ P_1^n\Big(L(X_1^n) \leq n\overline{\lambda}^{(n)}\Big) \leq -\frac{\varepsilon_1^2}{2\sigma_1^2}.$$

Question:

But, does the upper bound reflect the correct asymptotic scaling of this probability ?

Moderate-Deviations Analysis via the Martingale Approach (Cont.)

The following upper bound implies that

$$\lim_{n\to\infty} n^{1-2\eta} \ \ln \ P_1^n\Big(L(X_1^n) \leq n\overline{\lambda}^{(n)}\Big) \leq -\frac{\varepsilon_1^2}{2\sigma_1^2}.$$

Question:

But, does the upper bound reflect the correct asymptotic scaling of this probability ?

Reply:

Indeed, it follows from the moderate-deviations principle (MDP) for real-valued RVs.

Moderate-Deviations Principle (MDP) for Real-Valued RVs

Theorem

Let $\{X_i\}$ be i.i.d. random real-valued RVs, satisfying

$$\mathbb{E}[X_i] = 0, \quad \mathit{Var}(X_i) = \sigma^2, \quad |X_i| \leq d$$

and let $\eta \in (\frac{1}{2}, 1)$. Then, for every $\alpha > 0$,

$$\lim_{n\to\infty} n^{1-2\eta} \ln \mathbb{P}\left(\left| \sum_{i=1}^{n} X_i \right| \geq \alpha n^{\eta} \right) = -\frac{\alpha^2}{2\sigma^2}, \quad \forall\, \alpha \geq 0.$$

Moderate-Deviations Analysis via the Martingale Approach (Cont.)

- The MDP shows that this inequality holds in fact with equality.
- $\Rightarrow$ The refined inequality in Theorem 2 gives the correct asymptotic scaling in this case. It also gives an analytical bound for finite n.
- This is in contrast to the analysis which follows from the Azuma-Hoeffding inequality, which does not coincide with the correct asymptotic scaling (since $-\frac{\varepsilon_1^2}{2\sigma_1^2}$ is replaced by $-\frac{\varepsilon_1^2}{2d_1^2}$, and $\sigma_1 \leq d_1$).
- In the considered setting of moderate-deviations analysis for binary hypothesis testing, the error probability decays sub-exponentially to 0.

Papers on Moderate-Deviations Analysis in IT Aspects

1. E. A. Abbe, *Local to Global Geometric Methods in Information Theory*, Ph.D. dissertation, MIT, Boston, MA, USA, June 2008.
2. Y. Altuğ and A. B. Wagner, "Moderate-deviations analysis of channel coding: discrete memoryless case," Proc. ISIT 2010, pp. 265–269, Austin, Texas, USA, June 2010.
3. D. He, L. A. Lastras-Montaño, E. Yang, A. Jagmohan and J. Chen, "On the redundancy of Slepian-Wolf coding," *IEEE Trans. on Information Theory*, vol. 55, no. 12, pp. 5607–5627, December 2009.
4. Y. Polyanskiy and S. Verdú, "Channel dispersion and moderate-deviations limits of memoryless channels," Proceedings Forty-Eighth Annual Allerton Conference, pp. 1334–1339, UIUC, Illinois, USA, October 2010.
5. I. Sason, "Moderate-deviations analysis of binary hypothesis testing," `http://arxiv.org/abs/1111.1995`, November 2011.
6. V. Y. F. Tan, "Moderate-deviations of lossy source coding for discrete and Gaussian channels," `http://arxiv.org/abs/1111.2217`, November 2011.

Concentration Phenomena for Codes Defined on Graphs

Motivation & Background

- The performance analysis of a particular code is difficult, especially for codes of large block lengths.

Concentration Phenomena for Codes Defined on Graphs

Motivation & Background

- The performance analysis of a particular code is difficult, especially for codes of large block lengths.
- Does the performance concentrate around the average performance of the ensemble ?

Concentration Phenomena for Codes Defined on Graphs

Motivation & Background

- The performance analysis of a particular code is difficult, especially for codes of large block lengths.
- Does the performance concentrate around the average performance of the ensemble ?
- The existence of such a concentration validates the use of the density evolution technique as an analytical tool to assess performance of long enough codes (e.g., LDPC codes) and to assess their asymptotic gap to capacity.

Concentration Phenomena for Codes Defined on Graphs

Motivation & Background

- The performance analysis of a particular code is difficult, especially for codes of large block lengths.
- Does the performance concentrate around the average performance of the ensemble ?
- The existence of such a concentration validates the use of the density evolution technique as an analytical tool to assess performance of long enough codes (e.g., LDPC codes) and to assess their asymptotic gap to capacity.
- The current concentration results for codes defined on graphs, which mainly rely on the Azuma-Hoeffding inequality, are weak since in practice concentration is observed at much shorter block lengths.

Performance under Message-Passing Decoding

Theorem I - **[Concentration of performance under iterative message-passing decoding (Richardson and Urbanke, 2001)]**

Let $\mathcal{C}$, a code chosen uniformly at random from the ensemble LDPC(n, λ, ρ), be used for transmission over a memoryless binary-input output-symmetric (MBIOS) channel. Assume that the decoder performs l iterations of message-passing decoding, and let $P_{\mathsf{b}}(\mathcal{C}, l)$ denote the resulting bit error probability. Then, for every $\delta > 0$, there exists an $\alpha > 0$ where $\alpha = \alpha(\lambda, \rho, \delta, l)$ (*independent of the block length* n) such that

$$\mathbb{P}\left(|P_{\mathsf{b}}(\mathcal{C}, l) - \mathbb{E}_{\mathrm{LDPC}(n,\lambda,\rho)}[P_{\mathsf{b}}(\mathcal{C}, l)]| \geq \delta\right) \leq e^{-\alpha n}$$

Proof

The proof applies Azuma's inequality to a martingale sequence with bounded differences (IEEE Trans. on IT, Feb. 2001).

Conditional Entropy of LDPC code ensembles

Theorem II - **[Concentration of Conditional Entropy of LDPC code ensembles (Méasson et al. 2008)]**

Let $\mathcal{C}$ be chosen uniformly at random from the ensemble LDPC(n, λ, ρ). Assume that the transmission of the code $\mathcal{C}$ takes place over an MBIOS channel. Let $H(\mathbf{X}|\mathbf{Y})$ designate the conditional entropy of the transmitted codeword $\mathbf{X}$ given the received sequence $\mathbf{Y}$ from the channel. Then, for any $\xi > 0$,

$$\mathbb{P}\left(\left|H(\mathbf{X}|\mathbf{Y}) - \mathbb{E}_{\mathsf{LDPC}(n,\lambda,\rho)}[H(\mathbf{X}|\mathbf{Y})]\right| \geq \sqrt{n}\,\xi\right) \leq 2\exp(-B\xi^2)$$

where $B \triangleq \frac{1}{2(d_{\mathsf{c}}^{\max}+1)^2(1-R_{\mathsf{d}})}$, $d_{\mathsf{c}}^{\max}$ is the maximal check-node degree, and R_{d} is the design rate of the ensemble.

Proof - **[outline]**

1. Introduction of a martingale sequence with bounded differences:
   - Define the RV $Z = H_{\mathcal{G}}(\mathbf{X}|\mathbf{Y})$, where $\mathcal{G}$ is a graph of a code chosen uniformly at random from the ensemble LDPC(n, λ, ρ)
   - Define the martingale sequence $Z_t = \mathbb{E}[Z|\mathcal{F}_t] \quad t \in \{0, 1, \ldots, m\}$, where the filtration is the sequence of subsets of σ-algebras, generated by revealing each time another parity-check equation of the code.
2. Upper bounds on the differences $|Z_{t+1} - Z_t|$:
   - It was proved that $|Z_{t+1} - Z_t| \leq (r+1)\, H_{\mathcal{G}}(\tilde{X}|\mathbf{Y})$, where r is the degree of parity-check equation revealed at time t, and $\tilde{X} = X_{i_1} \oplus \ldots \oplus X_{i_r}$ (i.e., $\tilde{X}$ is the modulo-2 sum of some r bits in the codeword $\mathbf{X}$).
   - Then $r \leq d_{\mathsf{c}}^{\max}$, and $H_{\mathcal{G}}(\tilde{X}|\mathbf{Y}) \leq 1$.
3. Azuma's inequality was applied to a get a concentration inequality, using $|Z_{t+1} - Z_t| \leq d_{\mathsf{c}}^{\max} + 1$ for every $t = 0, \ldots, m-1$ where $m = n(1 - R_{\mathsf{d}})$ is the number of parity-check nodes.

Improvements to Theorem II

Improvement 1 - A tightened upper bound on the conditional entropy

- Instead of upper bounding $H_{\mathcal{G}}(\tilde{X}|\mathbf{Y})$ by 1, which is independent of the channel capacity (C), it can be proved that

$$H_{\mathcal{G}}(\tilde{X}|\mathbf{Y}) \leq h\left(\frac{1-C^{\frac{r}{2}}}{2}\right)$$

  where h is the binary entropy function to the base 2.
- For a BSC or BEC, this bound can be further improved.

Improvement 2 (trivial)

Instead of taking the trivial bound $r \leq d_{\mathsf{c}}^{\max}$ for all m terms in the Azuma's inequality, one can rely on the degree distribution of the parity-check nodes. The number of parity-check nodes of degree r is $n(1-R_{\mathsf{d}})\Gamma_r$.

Theorem III - [Tightened Expressions for B]

Considering the terms of Theorem II, applying these two improvements yields tightened expressions for B.

- General MBIOS - $B \triangleq \dfrac{1}{2(1-R_{\mathsf{d}})\sum_{i=1}^{d_{\mathsf{c}}^{\max}}(i+1)^2\Gamma_i\left[h\left(\frac{1-C^{\frac{i}{2}}}{2}\right)\right]^2}$
- BSC - $B \triangleq \dfrac{1}{2(1-R_{\mathsf{d}})\sum_{i=1}^{d_{\mathsf{c}}^{\max}}(i+1)^2\Gamma_i\left[h\left(\frac{1-[1-2h^{-1}(1-C)]^i}{2}\right)\right]^2}$
- BEC - $B \triangleq \dfrac{1}{2(1-R_{\mathsf{d}})\sum_{i=1}^{d_{\mathsf{c}}^{\max}}(i+1)^2\Gamma_i\,(1-C^i)^2}$

Numerical comparison for BEC and BIAWGN

Let us consider the case where

- $(2, 20)$ regular LDPC code ensemble.
- Communication over a BEC or BIAWGNC with capacity of 0.98 per channel use.

Compared to Theorems II, applying Theorem III results in tighter expressions for B:

- BIAWGN - Improvement by factor $\left[h\left(\frac{1-C^{\frac{d_c}{2}}}{2}\right)\right]^{-2} = 5.134$
- BEC - Improvement by factor $\frac{1}{\left(1-C^{d_c}\right)^2} = 9.051$

Comparison for Heavy-Tail Poisson Distribution (Tornado Codes)

Consider the capacity-achieving Tornado LDPC code ensemble for a BEC with erasure probability p. We wish to design a code ensemble that achieves a fraction $1-\varepsilon$ of the capacity.

- Theorem II - The parity-check degree is Poisson distributed, therefore $d_{\mathsf{c}}^{\max} = \infty$. Hence, $B = 0$ and this result is useless.
- Theorem III - B scales (at least) like $O\left(\frac{1}{\log^2\left(\frac{1}{\varepsilon}\right)}\right)$.

  The parameter B tends to zero slowly as we let the fractional gap ε tend to zero; this demonstrates a rather fast concentration.

Interim Conclusions

- The use of Azuma's inequality was addressed in the context of proving concentration phenomena for code ensembles defined on graphs and iterative decoding algorithms.
- A possible tightening of a concentration inequality for the conditional entropy of LDPC ensembles by using Azuma's inequality, and deriving an improved upper bound on the jumps of the martingale sequence.
- The improved inequality enables to prove concentration of the conditional entropy for ensembles of Tornado codes (in contrast to the original concentration inequality).

Concentration of Martingales in Coding Theory

Selected Papers Applying Azuma's Inequality for LDPC Ensembles

- M. Sipser and D. A. Spielman, "Expander codes," *IEEE Trans. on Information Theory*, vol. 42, no. 6, pp. 1710-1722, November 1996.
- M.G. Luby, M. Mitzenmacher, M.A. Shokrollahi and D.A. Spielman, "Efficient erasure correcting codes", *IEEE Trans. on Information Theory*, vol. 47, no. 2, pp. 569-584, February 2001.
- T. Richardson and R. Urbanke, "The capacity of low-density parity check codes under message-passing decoding." *IEEE Trans. on Information Theory*, vol. 47, pp. 599–618, February 2001.
- A. Kavcic, X. Ma and M. Mitzenmacher, "Binary intersymbol interference channels: Gallager bounds, density evolution, and code performance bounds," *IEEE Trans. on Information Theory*, vol. 49, no. 7, pp. 1636-1652, July 2003.

(Cont.)

- A. Montanari, "Tight bounds for LDPC and LDGM codes under MAP decoding," *IEEE Trans. on Information Theory*, vol. 51, no. 9, pp. 3247–3261, September 2005.
- C. Méasson, A. Montanari and R. Urbanke, "Maxwell construction: The hidden bridge between iterative and maximum a-posteriori decoding," *IEEE Trans. on Information Theory*, vol. 54, pp. 5277–5307, December 2008.
- L.R. Varshney, "Performance of LDPC codes under faulty iterative decoding," *IEEE Trans. on Information Theory*, vol. 57, no. 7, pp. 4427–4444, July 2011.

Orthogonal Frequency Division Multiplexing (OFDM)

- The OFDM modulation converts a high-rate data stream into a number of low-rate steams that are transmitted over parallel narrow-band channels.
- One of the problems of OFDM is that the peak amplitude of the signal can be significantly higher than the average amplitude.

  $\Rightarrow$ Sensitivity to non-linear devices in the communication path (e.g., digital-to-analog converters, mixers and high-power amplifiers).

  $\Rightarrow$ An increase in the symbol error rate and also a reduction in the power efficiency as compared to single-carrier systems.

OFDM (Cont.)

- Given an n-length codeword $\{X_i\}_{i=0}^{n-1}$, a single OFDM baseband symbol is described by

$$s(t; X_0, \ldots, X_{n-1}) = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} X_i \exp\Big(\frac{j\,2\pi i t}{T}\Big), \quad 0 \le t \le T.$$

- Assume that $X_0, \ldots, X_{n-1}$ are i.i.d. complex RVs with $|X_i| = 1$. Since the sub-carriers are orthonormal over $[0, T]$, then a.s. the power of the signal s over this interval is 1.
- The CF of the signal s, composed of n sub-carriers, is defined as

$$\mathsf{CF}_n(s) \triangleq \max_{0 \le t \le T} |s(t)|.$$

- The CF scales with high probability like $\sqrt{\log(n)}$ for large n.

Concentration of Measures

- In the following, we consider two of the main approaches for proving concentration inequalities, and apply them to prove concentration for the crest factor of OFDM signals.
  1. The 1st approach is based on martingales (the Azuma-Hoeffding inequality and some refinements).
  2. The 2nd approach is based on Talagrand's inequalities.

A Previously Reported Result

A concentration inequality for the CF of OFDM signals was derived (Litsyn and Wunder, IEEE Trans. on IT, 2006). It states that for every $c \geq 2.5$

$$\mathbb{P}\left(\left|\mathsf{CF}_n(s) - \sqrt{\log(n)}\right| < \frac{c \log\log(n)}{\sqrt{\log(n)}}\right) = 1 - O\left(\frac{1}{\left(\log(n)\right)^4}\right).$$

Theorem - [McDiarmid's Inequality]

- Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a vector of independent random variables with X_k taking values in a set A_k for each k.
- Suppose that a real-valued function f, defined on $\prod_k A_k$, satisfies

$$|f(\mathbf{x}) - f(\mathbf{x}')| \leq c_k$$

whenever the vectors $\mathbf{x}$ and $\mathbf{x}'$ differ only in the k-th coordinate.

- Let $\mu \triangleq \mathbb{E}[f(X)]$ be the expected value of $f(X)$.

Then, for every $\alpha \geq 0$,

$$\mathbb{P}(|f(X) - \mu| \geq \alpha) \leq 2 \exp\left(-\frac{2\alpha^2}{\sum_k c_k^2}\right).$$

Proving Concentration of the CF for OFDM Signals

Consider the case where $\{X_j\}_{j=0}^{n-1}$ are independent complex-valued random variables with magnitude 1, attaining the M points of an M-ary PSK constellation with equal probability.

Proving Concentration via the Azuma-Hoeffding Inequality

- Let us define

$$Y_i = \mathbb{E}[\,\mathsf{CF}_n(s)\,|\,X_0,\ldots,X_{i-1}], \quad i = 0,\ldots,n.$$

- $\{Y_i, \mathcal{F}_i\}_{i=0}^n$ is a martingale where $\mathcal{F}_i$ is the σ-algebra generated by $(X_0,\ldots,X_{i-1})$.
- This martingale has bounded jumps: $|Y_i - Y_{i-1}| \leq \frac{2}{\sqrt{n}}$ (revealing the i-th coordinate X_i affects the CF by at most $\frac{2}{\sqrt{n}}$).
- It follows from the Azuma-Hoeffding inequality that, for every $\alpha > 0$,

$$\mathbb{P}(|\mathsf{CF}_n(s) - \mathbb{E}[\mathsf{CF}_n(s)]| \geq \alpha) \leq 2\exp\left(-\frac{\alpha^2}{8}\right)$$

which demonstrates concentration around the expected value.

Proving Concentration via Martingales (cont.)

- The refined version of the Azuma-Hoeffding inequality improves the exponent by a factor of 2, due to the additional information on the conditional variance.

$$\mathbb{P}(|\mathsf{CF}_n(s) - \mathbb{E}[\mathsf{CF}_n(s)]| \geq \alpha) \leq 2\exp\left(-\frac{\alpha^2}{4}\left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right)\right).$$

- McDiarmid's inequality implies that, for every $\alpha \geq 0$,

$$\mathbb{P}(|\mathsf{CF}_n(s) - \mathbb{E}[\mathsf{CF}_n(s)]| \geq \alpha) \leq 2\exp\left(-\frac{\alpha^2}{2}\right)$$

  $\Rightarrow$ The exponent improves by a factor of 4 as compared to the Azuma-Hoeffding inequality.

- The same kind of result can be applied to QAM-modulated OFDM signals, since the independent RVs $\{X_j\}$ are bounded.

Talagrand's Inequality

- Talagrand's inequality is an approach used for establishing concentration results on product spaces, and this technique was introduced in Talagrand's landmark paper from 1995.
- We provide in the following two definitions that will be required for the introduction of a special form of Talagrand's inequalities.

Talagrand's Inequality (Cont.)

- Let $\mathbf{x}, \mathbf{y}$ be two n-length vectors. The Hamming distance between $\mathbf{x}$ and $\mathbf{y}$ is the number of coordinates where $\mathbf{x}$ and $\mathbf{y}$ disagree, i.e.,

$$d_{\mathsf{H}}(\mathbf{x}, \mathbf{y}) \triangleq \sum_{i=1}^{n} I_{\{x_i \neq y_i\}}$$

where I stands for the indicator function.

Generalization and normalization of the previous distance metric:

- Let $a = (a_1, \ldots, a_n) \in \mathbb{R}_+^n$ (i.e., a is a non-negative vector) satisfy $||a||_2 = 1$. Then, define

$$d_a(\mathbf{x}, \mathbf{y}) \triangleq \sum_{i=1}^{n} a_i I_{\{x_i \neq y_i\}}.$$

Hence, $d_{\mathsf{H}}(\mathbf{x}, \mathbf{y}) = \sqrt{n}\, d_a(\mathbf{x}, \mathbf{y})$ for $a = \left(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right)$.

Special Form of Talagrand's Inequality (Cont.)

Let the random vector $\mathbf{X} = (X_1, \ldots, X_n)$ be vector of independent random variables with X_k taking values in a set A_k, and let $A \triangleq \prod_{k=1}^{n} A_k$. Let $f : A \to \mathbb{R}$ satisfy the condition that, for every $\mathbf{x} \in A$, there exists a non-negative, normalized n-length vector $a = a(x)$ such that

$$f(\mathbf{x}) \leq f(\mathbf{y}) + \sigma d_a(\mathbf{x}, \mathbf{y}), \quad \forall\, \mathbf{y} \in A$$

for some fixed value $\sigma > 0$. Then, for every $\alpha \geq 0$,

$$\mathbb{P}(|f(X) - m| \geq \alpha) \leq 4 \exp\left(-\frac{\alpha^2}{4\sigma^2}\right)$$

where m is the median of $f(X)$
(i.e., $\mathbb{P}(f(X) \leq m) \geq \frac{1}{2}$ and $\mathbb{P}(f(X) \geq m) \geq \frac{1}{2}$).

Establishing Concentration via Talagrand's Inequality

- Let us assume that $X_0, Y_0, \ldots, X_{n-1}, Y_{n-1}$ are i.i.d. bounded complex RVs, and also for simplicity $|X_i| = |Y_i| = 1$.
- In order to apply Talagrand's inequality to prove concentration, note that

$$\begin{aligned}
&\max_{0\le t\le T}\big|\, s(t; X_0, \ldots, X_{n-1})\big| - \max_{0\le t\le T}\big|\, s(t; Y_0, \ldots, Y_{n-1})\big| \\
&\le \max_{0\le t\le T}\big|\, s(t; X_0, \ldots, X_{n-1}) - s(t; Y_0, \ldots, Y_{n-1})\big| \\
&\le \frac{1}{\sqrt{n}} \left| \sum_{i=0}^{n-1} (X_i - Y_i) \exp\Big(\frac{j\,2\pi i t}{T}\Big) \right| \\
&\le \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} |X_i - Y_i| \\
&\le \frac{2}{\sqrt{n}} \sum_{i=0}^{n-1} I_{\{x_i \neq y_i\}}
\end{aligned}$$

Establishing Concentration via Talagrand's Inequality (Cont.)

- Talagrand's inequality implies that, for every $\alpha \geq 0$,

$$\mathbb{P}(|\mathsf{CF}_n(s) - m_n| \geq \alpha) \leq 4\exp\left(-\frac{\alpha^2}{16}\right), \quad \forall\, \alpha > 0$$

  where m_n is the median of the crest factor for OFDM signals that are composed of n sub-carriers.
- This inequality demonstrates the concentration of this measure around its median.

Establishing Concentration via Talagrand's Inequality (Cont.)

Corollary

The median and expected value of the crest factor differ by at most a constant, independently of the number of sub-carriers n.

Proof: From Talagrand's inequality

$$\begin{aligned}
&|\mathbb{E}[\mathsf{CF}_n(s)] - m_n| \\
&\leq \mathbb{E}\,|\mathsf{CF}_n(s) - m_n| \\
&= \int_0^\infty \mathbb{P}(|\mathsf{CF}_n(s) - m_n| \geq \alpha)\, d\alpha \leq 8\sqrt{\pi}.
\end{aligned}$$

where the equality holds since for a non-negative random variable Z

$$\mathbb{E}[Z] = \int_0^\infty \mathbb{P}(Z \geq t)\, dt.$$

Achievable Rates of Random Codes in Nonlinear Channels

Non-Linear Channels

Non-linear effects are typically encountered in wireless and optical communication systems:

- Traveling-wave tube amplifiers (TWTA) on board satellites operate at or near the saturation region to obtain high power efficiency.
- Non-linearities in optical fibers.

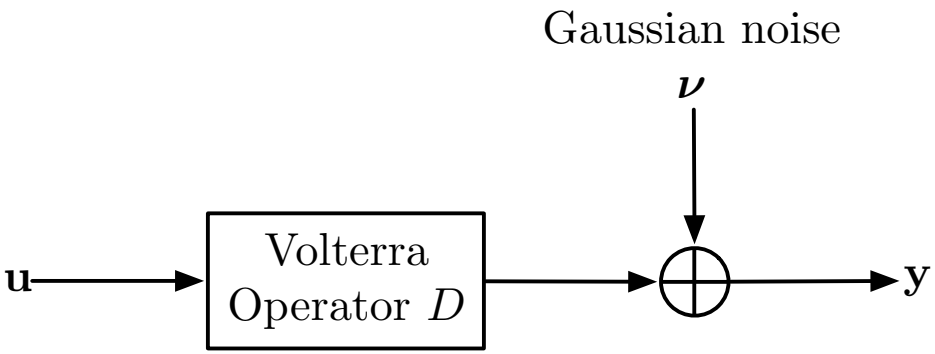
Non-linear effects $\Rightarrow$ Degradation of the quality of info. transmission.

Nonlinear Volterra channels

- Input-output model: $y_i = [D\mathbf{u}]_i + \nu_i$ (i is the time index).
- Volterra's operator D of order L and memory q:

$$[D\mathbf{u}]_i = h_0 + \sum_{j=1}^{L} \sum_{i_1=0}^{q} \cdots \sum_{i_j=0}^{q} h_j(i_1, \ldots, i_j) u_{i-i_1} \ldots u_{i-i_j}.$$

- $\boldsymbol{\nu}$ is an additive Gaussian noise vector with i.i.d. components $\nu_i \sim \mathcal{N}(0, \sigma_\nu^2)$.



Goal

Establishing new achievable rates for non-linear Volterra communication channels, and exemplifying the characteristics of these rates.

In this part of the talk

- New achievable rates for nonlinear Volterra channels are derived.
- The approach relies on exponential martingale inequalities that form some refined versions of the Azuma-Hoeffding inequality.
- The bounds are applied to linear channels with or without memory, memoryless nonlinear channels, and Volterra (non-linear) models.

Achievable Rates of Random Coding under ML Decoding

Setting

- Consider an ensemble of block codes $\mathbf{C}$ of length N and rate R.
- Let $\mathcal{C} \in \mathbf{C}$ be a codebook in the ensemble. The number of codewords in $\mathcal{C}$ is $M = \lceil \exp(NR) \rceil$.
- The codewords of a codebook $\mathcal{C}$ are assumed to be independent, and the symbols in each codeword are assumed to be i.i.d. with an arbitrary probability distribution P.
- An ML decoding algorithm is assumed.

Achievable Rates of Random Coding (Cont.)

Analysis

- An ML decoding error occurs if, given the transmitted message m and the received vector $\mathbf{y}$, there exists another message $m' \neq m$ such that

$$||\mathbf{y} - D\mathbf{u}_{m'}||_2 \leq ||\mathbf{y} - D\mathbf{u}_m||_2.$$

- The union bound for an AWGN channel implies that

$$P_{\mathsf{e}|m}(\mathcal{C}) \leq \sum_{m' \neq m} Q\left(\frac{\|D\mathbf{u}_m - D\mathbf{u}_{m'}\|_2}{2\sigma_\nu}\right)$$

where

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t, \quad \forall\, x \in$$

is the complementary Gaussian cumulative distribution function.

Achievable Rates of Random Coding (Cont.)

Analysis (Cont.)

- Since $Q(x) \leq \frac{1}{2}\exp\left(-\frac{x^2}{2}\right)$ for $x \geq 0$, then for $\rho \in [0,1]$

$$P_{\mathsf{e}|m}(\mathcal{C}) \leq \sum_{m' \neq m} \exp\left(-\frac{\rho \left\| D\mathbf{u}_m - D\mathbf{u}_{m'} \right\|_2^2}{8\sigma_\nu^2}\right).$$

  At this stage, the optimal value is $\rho_{\mathsf{opt}} = 1$.

- The average ML decoding error probability over the ensemble satisfies

$$\overline{P}_{\mathsf{e}|m} \leq \mathbb{E}\left[\sum_{m' \neq m} \exp\left(-\frac{\rho \left\| D\mathbf{u}_m - D\mathbf{u}_{m'} \right\|_2^2}{8\sigma_\nu^2}\right)\right]$$

Achievable Rates of Random Coding (Cont.)

Analysis (Cont.)

- The average ML decoding error probability over the code ensemble and the transmitted message satisfies, for an arbitrary $\rho \in [0,1]$,

$$\overline{P}_{\mathsf{e}} \leq (M-1)\,\mathbb{E}\left[\exp\left(-\frac{\rho\,\|D\mathbf{u}-D\widetilde{\mathbf{u}}\|_2^2}{8\sigma_\nu^2}\right)\right]$$

  where the expectation is taken over two randomly chosen codewords $\mathbf{u}$ and $\widetilde{\mathbf{u}}$ where these codewords are independent, and their symbols are i.i.d. with a probability distribution P.

- Consider a filtration $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}_N$ where $\mathcal{F}_i$ is given by

$$\mathcal{F}_i \triangleq \sigma(U_1, \widetilde{U}_1, \ldots, U_i, \widetilde{U}_i), \quad \forall\, i \in \{1,\ldots,N\}$$

  for two randomly selected codewords $\mathbf{u} = (u_1,\ldots,u_N)$, and $\widetilde{\mathbf{u}} = (\tilde{u}_1,\ldots,\tilde{u}_N)$ from the codebook.

Achievable Rates of Random Coding (Cont.)

Analysis (Cont.)

- $\mathcal{F}_i$ is the minimal σ-algebra that is generated by the first i coordinates of these two codewords.
- Define the discrete-time martingale $\{X_k, \mathcal{F}_k\}_{k=0}^{N}$ by

$$X_k = \mathbb{E}[||D\mathbf{u} - D\widetilde{\mathbf{u}}||_2^2 \,|\, \mathcal{F}_k]$$

  designates the conditional expectation of the squared Euclidean distance between the distorted codewords $D\mathbf{u}$ and $D\widetilde{\mathbf{u}}$ given the first i coordinates of the two codewords $\mathbf{u}$ and $\widetilde{\mathbf{u}}$.
- The first and last elements of this martingale sequence are, respectively, equal to

$$X_0 = \mathbb{E}\,[||D\mathbf{u} - D\widetilde{\mathbf{u}}||_2^2], \quad X_N = ||D\mathbf{u} - D\widetilde{\mathbf{u}}||_2^2.$$

Achievable Rates of Random Coding (Cont.)

Analysis (Cont.)

- Let $\xi_k = X_k - X_{k-1}$ be the jumps of the martingale, then

$$\sum_{k=1}^{N} \xi_k = X_N - X_0 = ||D\mathbf{u} - D\widetilde{\mathbf{u}}||_2^2 - \mathbb{E}\left[||D\mathbf{u} - D\widetilde{\mathbf{u}}||_2^2\right].$$

- Hence, substitution in the bound two slides earlier gives

$$\overline{P}_{\mathsf{e}} \leq \exp(NR)\ \exp\left(-\frac{\rho\,\mathbb{E}\left[||D\mathbf{u} - D\widetilde{\mathbf{u}}||_2^2\right]}{8\sigma_\nu^2}\right)\mathbb{E}\left[\exp\left(-\frac{\rho}{8\sigma^2}\cdot\sum_{k=1}^{N}\xi_k\right)\right].$$

Achievable Rates of Random Coding (Cont.)

Analysis (Cont.)

- Since the codewords are independent and their symbols are i.i.d., then it follows that

$$\mathbb{E}||D\mathbf{u} - D\widetilde{\mathbf{u}}||_2^2 = 2\left(\sum_{k=1}^{q-1} \mathrm{Var}\left([D\mathbf{u}]_k\right) + \sum_{k=q}^{N} \mathrm{Var}\left([D\mathbf{u}]_k\right)\right).$$

- Due to the channel model and the assumption that the symbols $\{u_i\}$ are i.i.d., it follows that $\mathrm{Var}\left([D\mathbf{u}]_k\right)$ is fixed for $k = q, \ldots, N$.
- Let $D_{\mathsf{v}}(P)$ designate this common value of the variance (i.e., $D_{\mathsf{v}}(P) = \mathrm{Var}\left([D\mathbf{u}]_k\right)$ for $k \geq q$), then

$$\mathbb{E}||D\mathbf{u} - D\widetilde{\mathbf{u}}||_2^2 = 2\left(\sum_{k=1}^{q-1} \mathrm{Var}\left([D\mathbf{u}]_k\right) + (N-q+1)D_{\mathsf{v}}(P)\right).$$

Achievable Rates of Random Coding (Cont.)

Analysis (Cont.)

- Assume that $||\mathbf{u}||_\infty \leq K < +\infty$ holds a.s. for some $K > 0$, and it is independent of the block length N.
- It implies that, for some finite constant $C(P)$ and for every $\rho \in [0,1]$

$$\overline{P}_{\mathsf{e}} \leq C_\rho(P)\ \exp\left\{-N\left(\frac{\rho\, D_{\mathsf{v}}(P)}{4\sigma_\nu^2} - R\right)\right\}\ \mathbb{E}\left[\exp\left(\frac{\rho}{8\sigma_\nu^2}\cdot\sum_{k=1}^{N} Z_k\right)\right],$$

where $Z_k \stackrel{\triangle}{=} -\xi_k$, so $\{Z_k, \mathcal{F}_k\}$ is a martingale-difference that corresponds to the jumps of the martingale $\{-X_k, \mathcal{F}_k\}$.

$$\begin{aligned} Z_k &= X_{k-1} - X_k \\ &= \mathbb{E}[||D\mathbf{u} - D\widetilde{\mathbf{u}}||_2^2 \mid \mathcal{F}_{k-1}] - \mathbb{E}[||D\mathbf{u} - D\widetilde{\mathbf{u}}||_2^2 \mid \mathcal{F}_k]. \end{aligned}$$

Martingale Inequalities

Theorem 1

Let $\{X_k, \mathcal{F}_k\}_{k=0}^n$, for some $n \in \mathbb{N}$, be a discrete-parameter, real-valued martingale with bounded jumps. Let

$$\xi_k \triangleq X_k - X_{k-1}, \quad \forall\, k \in \{1, \ldots, n\}$$

designate the jumps of the martingale. Assume that, for some constants $d, \sigma > 0$, the following two requirements

$$\xi_k \leq d, \quad \mathsf{Var}(\xi_k | \mathcal{F}_{k-1}) \leq \sigma^2$$

hold almost surely (a.s.) for every $k \in \{1, \ldots, n\}$. Let $\gamma \triangleq \frac{\sigma^2}{d^2}$. Then, for every $t \geq 0$,

$$\mathbb{E}\left[\exp\left(t \sum_{k=1}^n \xi_k\right)\right] \leq \left(\frac{e^{-\gamma t d} + \gamma e^{t d}}{1+\gamma}\right)^n .$$

Martingale Inequalities (Cont.)

Theorem 2

Let $\{X_k, \mathcal{F}_k\}_{k=0}^n$, for some $n \in \mathbb{N}$, be a discrete-time, real-valued martingale with bounded jumps. Let $\xi_k \triangleq X_k - X_{k-1}, \quad \forall\, k \in \{1, \ldots, n\}$ and let $m \in \mathbb{N}$ be an even number, $d > 0$ be a positive number, and $\{\mu_l\}_{l=2}^m$ be a sequence of numbers such that

$$\xi_k \leq d, \qquad \mathbb{E}\big[(\xi_k)^l \,|\, \mathcal{F}_{k-1}\big] \leq \mu_l, \quad \forall\, l \in \{2, \ldots, m\}$$

holds a.s. for every $k \in \{1, \ldots, n\}$. Furthermore, let $\gamma_l \triangleq \frac{\mu_l}{d^l}$ for every $l \in \{2, \ldots, m\}$. Then, for every $t \geq 0$,

$$\mathbb{E}\left[\exp\left(t \sum_{k=1}^n \xi_k\right)\right] \leq \left(1 + \sum_{l=2}^{m-1} \frac{(\gamma_l - \gamma_m)\,(td)^l}{l!} + \gamma_m (e^{td} - 1 - td)\right)^n .$$

Achievable Rates for Random Coding

First Bounding Technique

The maximal achievable rate for random coding, which follows from the union bound and Theorem 1, is given by

$$R_1(\sigma_\nu^2) = \max_P \begin{cases} D\left(\left(\frac{\gamma_2}{1+\gamma_2} + \frac{2\mathcal{D}_{\mathsf{v}}(P)}{d(1+\gamma_2)}\right) \,||\, \frac{\gamma_2}{1+\gamma_2}\right), & \text{if } D_{\mathsf{v}}(P) < \frac{\gamma_2\, d\left(\exp\left(\frac{d(1+\gamma_2)}{8\sigma_\nu^2}\right)-1\right)}{2\left(1+\gamma_2\exp\left(\frac{d(1+\gamma_2)}{8\sigma_\nu^2}\right)\right)} \\ \frac{D_{\mathsf{v}}(P)}{4\sigma_\nu^2} - \ln\left(\frac{\exp\left(-\frac{\gamma_2\, d}{8\sigma_\nu^2}\right)+\gamma_2\exp\left(\frac{d}{8\sigma_\nu^2}\right)}{1+\gamma_2}\right), & \text{otherwise} \end{cases}$$

where

$$D(p||q) \triangleq p\ln\left(\frac{p}{q}\right) + (1-p)\ln\left(\frac{1-p}{1-q}\right), \quad \forall\, p,q \in (0,1).$$

Achievable Rates for Random Coding

Second Bounding Technique

- Assume that for some even number $m \in \mathbb{N}$

$$Z_k \leq d, \qquad \mathbb{E}\big[(Z_k)^l \,|\, \mathcal{F}_{k-1}\big] \leq \mu_l, \;\; \forall\, l \in \{2, \ldots, m\}$$

  hold a.s. for some positive constant $d > 0$ and a sequence $\{\mu_l\}_{l=2}^m$.
- Let $\gamma_l \triangleq \frac{\mu_l}{d^l}$ for every $l \in \{2, \ldots, m\}$.
- The maximal achievable rate that follows from Theorem 2 is given by

$$R_2(\sigma_\nu^2) \triangleq \max_{P} \max_{\rho \in [0,1]} \left\{ \frac{\rho\, D_{\mathsf{v}}(P)}{4\sigma_\nu^2} - \ln\left(1 + \sum_{l=2}^{m-1} \frac{\gamma_l - \gamma_m}{l!} \left(\frac{\rho d}{8\sigma_\nu^2}\right)^l + \gamma_m \left(\exp\left(\frac{\rho\, d}{8\,\sigma_\nu^2}\right) - 1 - \frac{\rho\, d}{8\,\sigma_\nu^2}\right)\right)\right\}.$$

Achievable Rates for Random Coding over Specific Channel Models

Example: Binary-Input AWGN Channel

- Consider the case of a binary-input AWGN channel where

$$Y_k = U_k + \nu_k$$

where $U_i = \pm A$ for some constant $A > 0$ is a binary input, and $\nu_i \sim \mathcal{N}(0, \sigma_\nu^2)$ is an additive Gaussian noise with zero mean and variance σ_ν^2.

- Since the codewords $\mathbf{U} = (U_1, \ldots, U_N)$ and $\widetilde{\mathbf{U}} = (\widetilde{U}_1, \ldots, \widetilde{U}_N)$ are independent and their symbols are i.i.d., let for some $\alpha \in [0, 1]$

$$P(U_k = A) = P(\widetilde{U}_k = A) = \alpha,$$
$$P(U_k = -A) = P(\widetilde{U}_k = -A) = 1 - \alpha.$$

Achievable Rates for Random Coding: Examples

Example: Binary-Input AWGN Channel (Cont.)

- Since the channel is memoryless and the all the symbols are i.i.d.

$$\begin{aligned} Z_k &= \mathbb{E}[||\mathbf{U}-\widetilde{\mathbf{U}}||_2^2 \,|\, \mathcal{F}_{k-1}] - \mathbb{E}[||\mathbf{U}-\widetilde{\mathbf{U}}||_2^2 \,|\, \mathcal{F}_k] \\ &= 8\alpha(1-\alpha)A^2 - (U_k - \widetilde{U}_k)^2. \end{aligned}$$

- $\Rightarrow$ For every k, $Z_k \leq 8\alpha(1-\alpha)A^2 \triangleq d$, and for every $k, l \in \mathbb{N}$

$$\begin{aligned} &\mathbb{E}\left[(Z_k)^l \,|\, \mathcal{F}_{k-1}\right] \\ &= \left[1-2\alpha(1-\alpha)\right]\left(8\alpha(1-\alpha)A^2\right)^l + 2\alpha(1-\alpha)\left(8\alpha(1-\alpha)A^2 - 4A^2\right)^l \\ &\triangleq \mu_l. \end{aligned}$$

- $\Rightarrow \gamma_l \triangleq \frac{\mu_l}{d^l} = \left[1-2\alpha(1-\alpha)\right]\left[1+(-1)^l\left(\frac{1-2\alpha(1-\alpha)}{2\alpha(1-\alpha)}\right)^{l-1}\right].$

Achievable Rates for Random Coding: Examples

Example: Binary-Input AWGN Channel (Cont.)

- Assume that the binary input is symmetric, so $\alpha = \frac{1}{2}$ and $P = (\frac{1}{2}, \frac{1}{2})$.
- In this case,

$$D_{\mathsf{v}}(P) = \mathsf{Var}(U_k) = A^2, \quad d = 2A^2, \qquad \gamma_l = \frac{1 + (-1)^l}{2}, \ \ \forall\, l \in \mathbb{N}.$$

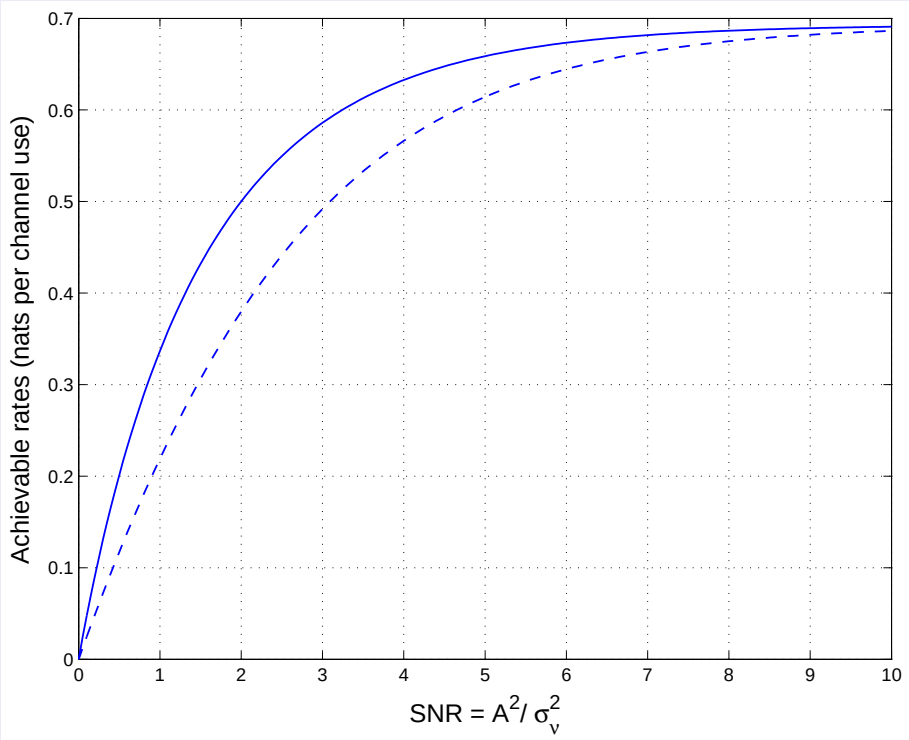
- In this case, the first and second achievable rates coincide when one takes $m \to \infty$ (with m even).
- The common value in this case is

$$R_1(\mathsf{SNR}) = \frac{\mathsf{SNR}}{4} - \ln \cosh\left(\frac{\mathsf{SNR}}{4}\right)$$

in units of nats per channel use where $\mathsf{SNR} \triangleq \frac{A^2}{\sigma_\nu^2}$ designates the signal to noise ratio.

Achievable Rates for Random Coding: Examples

Example: Binary-Input AWGN Channel (Cont.)



A comparison between the symmetric i.i.d. mutual information of the binary-input AWGN channel (solid line) and the achievable rate (dashed line) that follows from the martingale approach & the union bound.

Achievable Rates for Random Coding: Examples

A 3-rd order Volterra channel

Kernels of a 3rd order Volterra system with memory 2

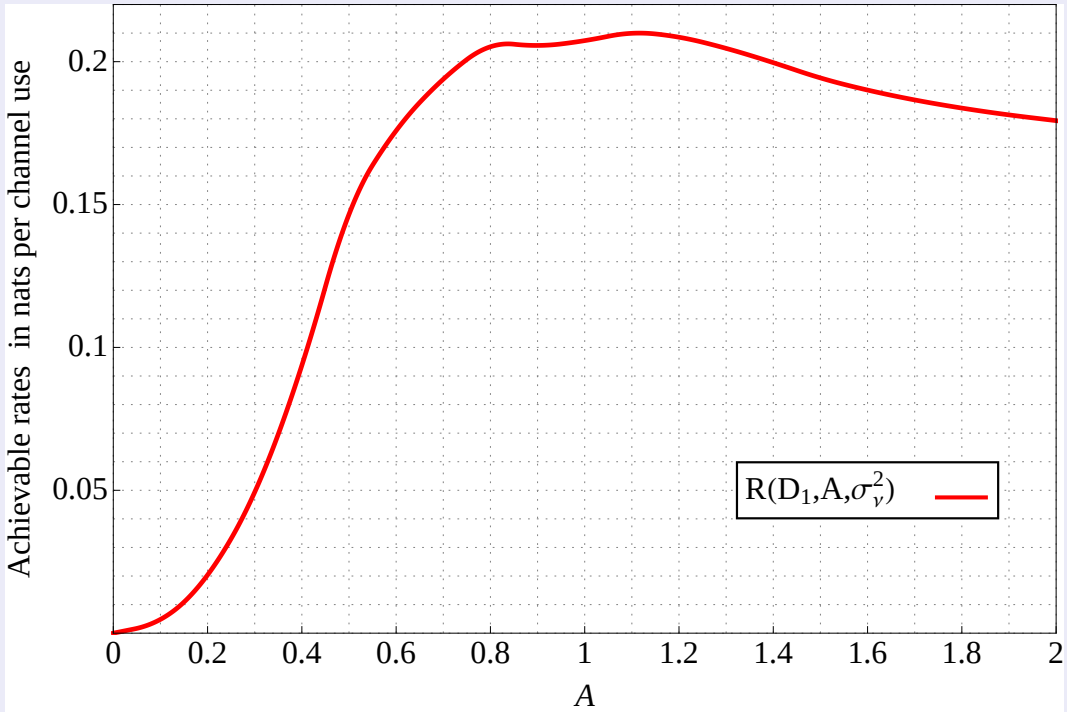
kernel	$h_1(0)$	$h_1(1)$	$h_1(2)$	$h_2(0,0)$	$h_2(1,1)$	$h_2(0,1)$
value	1.0	0.5	-0.8	1.0	-0.3	0.6

kernel	$h_3(0,0,0)$	$h_3(1,1,1)$	$h_3(0,0,1)$	$h_3(0,1,1)$	$h_3(0,1,2)$
value	1.0	-0.5	1.2	0.8	0.6

- $h_j(i_1, i_2, \ldots, i_j)$ coefficient of $u_{i-i_1} u_{i-i_2} \ldots u_{i-i_j}$
  Example: $h_1(0) \rightarrow u_i$, $h_3(0,0,1) \rightarrow u_i^2 u_{i-1}$
- Analytic calculation of the martingale parameters d and γ_2 when the input channel is binary are done to obtain achievable rates for random coding.

Achievable Rates for Random Coding: Examples

A 3-rd order Volterra channel - Achievable rates



Ongoing work (jointly with K. Xenoulis and N. Kalouptsidis)

- Improvements in the low SNR regime via existing improvements to Bennett's inequality (`http://arxiv.org/abs/1206.2592`).
- For time-invariant ISI channels, it is possible to calculate the parameters $\{\gamma_l\}_{l\geq 2}$ and d, so one needs to compare the achievable rates via the martingale approach with some existing bounds.
- Using better bounds than the union bound to possibly get improved achievable rates.

The reason for the weakness of the bounds at low SNR is mainly attributed to use of the union bound.

References

1. S. Benedetto and E. Biglieri, *Principles of Digital Transmission with Wireless Applications*, Kluwer Academic/ Plenum Publishers, 1999.
2. S. Boyd, L. O. Chua and C. A. Desoer, "Analytical foundations of Volterra series," *IMA Journal of Mathematical Control & Information*, vol. 1, pp. 243–282, 1984.
3. C. McDiarmid, "Concentration," *Probabilistic Methods for Algorithmic Discrete Mathematics*, pp. 195–248, Springer, 1998.
4. K. Xenoulis and N. Kalouptsidis, "Achievable rates for nonlinear Volterra channels," *IEEE Trans. on Information Theory*, vol. 57, no. 3, pp. 1237–1248, March 2011.
5. K. Xenoulis, N. Kalouptsidis and I. Sason, "New achievable rates for nonlinear Volterra channels via martingale inequalities," *Proceedings of the 2012 IEEE International Symposium on Information Theory (ISIT 2012)*, pp. 1430–1434, MIT, Boston, MA, USA, July 2012.

Summary and Conclusions

- This talk especially focused on concentration inequalities for martingales, followed by some of their applications and implications in information theory, communications and coding.

Summary and Conclusions

- This talk especially focused on concentration inequalities for martingales, followed by some of their applications and implications in information theory, communications and coding.
- The area of concentration of measure has seen enormous growth and tremendous activity since the early '90s.

Summary and Conclusions

- This talk especially focused on concentration inequalities for martingales, followed by some of their applications and implications in information theory, communications and coding.
- The area of concentration of measure has seen enormous growth and tremendous activity since the early '90s.
- Several approaches, that are used to derive concentration inequalities, have been well assimilated into the culture of probability theory, extending and generalizing results in various different directions.

Summary and Conclusions

- This talk especially focused on concentration inequalities for martingales, followed by some of their applications and implications in information theory, communications and coding.
- The area of concentration of measure has seen enormous growth and tremendous activity since the early '90s.
- Several approaches, that are used to derive concentration inequalities, have been well assimilated into the culture of probability theory, extending and generalizing results in various different directions.
- These probabilistic results are useful to establish results of theoretical and practical interest in computer science, machine learning, information theory and statistics, communications, coding theory, statistical physics and geometry.