On Concentration for Martingales and Applications in Information Theory and Communications

Igal Sason

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

Information Theory and Applications (ITA) Workshop
San Diego, California, USA

February 5–10, 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.

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 ($\mathbb{E}[X]$).

Application: Iterative 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

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

Proof

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

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.*

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

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

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

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

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

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

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

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

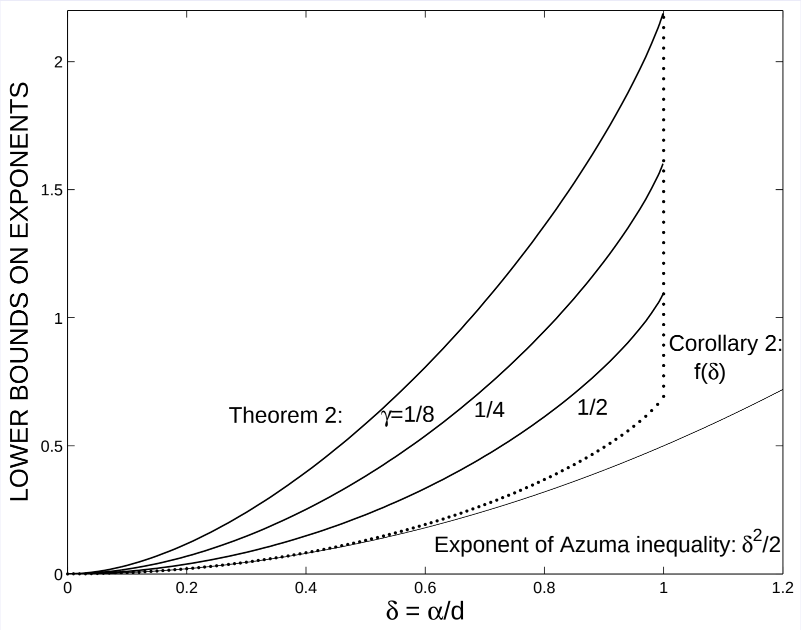
and $h_2(x) \triangleq -x \log_2(x) - (1-x)\log_2(1-x)$ for $0 \leq x \leq 1$.

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

More on Refined Versions of the Azuma-Hoeffding Inequality

- For concentration inequalities, dependent on some higher conditional centered moments of X_k given $\mathcal{F}_{k-1}$ where a.s.

$$\mathbb{E}\big[(X_k - X_{k-1})^l \,|\, \mathcal{F}_{k-1}\big] \leq \mu_l, \quad \forall\, l \geq 2$$

(which are finite since $|\xi_k| \triangleq |X_k - X_{k-1}| \leq d$ a.s.), we introduce Theorem 3 that was derived in `http://arxiv.org/abs/1111.1977`.

More on Refined Versions of the Azuma-Hoeffding Inequality

- For concentration inequalities, dependent on some higher conditional centered moments of X_k given $\mathcal{F}_{k-1}$ where a.s.

$$\mathbb{E}\big[(X_k - X_{k-1})^l \,|\, \mathcal{F}_{k-1}\big] \leq \mu_l, \quad \forall\, l \geq 2$$

  (which are finite since $|\xi_k| \triangleq |X_k - X_{k-1}| \leq d$ a.s.), we introduce Theorem 3 that was derived in `http://arxiv.org/abs/1111.1977`.
- These concentration inequalities can be adapted to sub and super martingales (as shown in Section III of the same paper).

Theorem (I. S., 2011)

Let $\{X_k, \mathcal{F}_k\}_{k=0}^{\infty}$ *be a real-valued martingale, and let* $m \in \mathbb{N}$ *be an even number. Assume that the following conditions hold a.s. for every* $k \in \mathbb{N}$

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

for some $d > 0$ *and non-negative numbers* $\{\mu_l\}_{l=2}^{m}$*. Then, for every* $\alpha \geq 0$,

$$\mathbb{P}(|X_n - X_0| \geq n\alpha) \leq 2 \left\{ \min_{x \geq 0} e^{-\delta x} \left[1 + \sum_{l=2}^{m-1} \frac{(\gamma_l - \gamma_m) x^l}{l!} + \gamma_m (e^x - 1 - x) \right] \right\}^n$$

where

$$\delta \triangleq \frac{\alpha}{d}, \quad \gamma_l \triangleq \frac{\mu_l}{d^l}, \quad \forall\, l = 2, \ldots, m.$$

Remark

The concentration inequality in Theorem 3 improves by increasing the value of m, and its exponential scaling converges fast in terms of m. There is a necessary and sufficient condition that implies that Theorem 3 gives a sharper result then Theorem 2.

Relations 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.

Applications in Information Theory, Communications and Coding

- Revisited Large deviations analysis of binary hypothesis testing (http://arxiv.org/abs/1111.1992).

Applications in Information Theory, Communications and Coding

- Revisited Large deviations analysis of binary hypothesis testing (http://arxiv.org/abs/1111.1992).
- Moderate deviations analysis of binary hypothesis testing (http://arxiv.org/abs/1111.1995, submitted to ISIT 2012).

Applications in Information Theory, Communications and Coding

- Revisited Large deviations analysis of binary hypothesis testing (http://arxiv.org/abs/1111.1992).
- Moderate deviations analysis of binary hypothesis testing (http://arxiv.org/abs/1111.1995, submitted to ISIT 2012).
- Concentration results for the peak-to-average power ratio of OFDM signals (http://arxiv.org/abs/1111.1982 & presented at ISWCS '11, Aachen, Germany, Nov. '11).

Applications in Information Theory, Communications and Coding

- Revisited Large deviations analysis of binary hypothesis testing (http://arxiv.org/abs/1111.1992).
- Moderate deviations analysis of binary hypothesis testing (http://arxiv.org/abs/1111.1995, submitted to ISIT 2012).
- Concentration results for the peak-to-average power ratio of OFDM signals (http://arxiv.org/abs/1111.1982 & presented at ISWCS '11, Aachen, Germany, Nov. '11).
- Concentration results for LDPC code ensembles (ISIT 2011, joint work with Ronen Eshel. For more details, see his Master thesis: R. Eshel, Aspects of Convex Optimization and Concentration in Coding, Technion, Haifa, Israel, February 2012).

Applications in Information Theory, Communications and Coding

- Revisited Large deviations analysis of binary hypothesis testing (http://arxiv.org/abs/1111.1992).
- Moderate deviations analysis of binary hypothesis testing (http://arxiv.org/abs/1111.1995, submitted to ISIT 2012).
- Concentration results for the peak-to-average power ratio of OFDM signals (http://arxiv.org/abs/1111.1982 & presented at ISWCS '11, Aachen, Germany, Nov. '11).
- Concentration results for LDPC code ensembles (ISIT 2011, joint work with Ronen Eshel. For more details, see his Master thesis: R. Eshel, Aspects of Convex Optimization and Concentration in Coding, Technion, Haifa, Israel, February 2012).
- New achievable rates for ISI and nonlinear Volterra channels via the martingale approach (joint work of Kostis Xenoulis and Nicholas Kalouptsidis, IEEE Trans. on IT, March 2011).

Summary and Conclusions

- This talk focused on refined versions of the Azuma-Hoeffding inequality for martingales, followed by some of their implications and applications in information theory, communications and coding.
- The area of measure concentration 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.
- Theorem 3 is a new concentration inequality, derived via the martingale approach.
- Some more material related to this talk is available in my home page (a line of work in progress).