On Concentration and Moderate-Deviations Analysis of Binary Hypothesis Testing

Igal Sason

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

ISIT 2012, MIT, Boston, MA, USA
July 2012.

Outline of the Talk

In this talk, we

- Introduce a known concentration inequality which is a refinement of the Azuma-Hoeffding inequality for real-valued discrete-parameter martingales.

Outline of the Talk

In this talk, we

- Introduce a known concentration inequality which is a refinement of the Azuma-Hoeffding inequality for real-valued discrete-parameter martingales.
- This inequality will be applied to the moderate-deviations analysis of binary hypothesis testing.

Outline of the Talk

In this talk, we

- Introduce a known concentration inequality which is a refinement of the Azuma-Hoeffding inequality for real-valued discrete-parameter martingales.
- This inequality will be applied to the moderate-deviations analysis of binary hypothesis testing.
- Discuss the result in light of some related work.

Outline of the Talk

In this talk, we

- Introduce a known concentration inequality which is a refinement of the Azuma-Hoeffding inequality for real-valued discrete-parameter martingales.
- This inequality will be applied to the moderate-deviations analysis of binary hypothesis testing.
- Discuss the result in light of some related work.
- This talk partially presents results from a recent work:

  I. Sason, "On Refined Versions of the Azuma-Hoeffding Inequality with Applications in Information Theory," available at: `http://arxiv.org/abs/1111.1977`.

Concentration via the Martingale Approach

Theorem - [Azuma-Hoeffding inequality]

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

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

then

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

Theorem (Th. 2 – Refined Version I of Azuma-Hoeffding Inequality)

Let $\{X_k, \mathcal{F}_k\}_{k=0}^n$ *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* $D(p||q) \triangleq p \ln\left(\frac{p}{q}\right) + (1 - p) \ln\left(\frac{1-p}{1-q}\right)$.

Proposition

Let $\{X_k, \mathcal{F}_k\}_{k=0}^{\infty}$ *be a real-valued martingale. Then, for every* $\alpha \geq 0$,

$$\mathbb{P}(|X_n - X_0| \geq \alpha\sqrt{n}) \leq 2\exp\Big(-\frac{\delta^2}{2\gamma}\Big)\Big(1 + O\big(n^{-\frac{1}{2}}\big)\Big).$$

Proof

This follows from Theorem 2 by replacing δ with $\delta' \triangleq \frac{\delta}{\sqrt{n}}$.

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

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

and

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

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

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

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

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

Moderate-Deviations Analysis via the Martingale Approach

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

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

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

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

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

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 $\{X_i\}$

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

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

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

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

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

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

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

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

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

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 $\sigma_1 \le d_1$).
- 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.