Probabilistic Methods in Combinatorics
Lecture 10

Linyuan Lu

University of South Carolina

Mathematical Sciences Center at Tsinghua University

November 16, 2011 – December 30, 2011

Martingale

A martingale is a sequence $X_0, X_1, \ldots, X_m$ of random variables so that for $0 \le i < m$,

$$E(X_{i+1} | X_i, \ldots, X_0) = X_i.$$

Martingale

A martingale is a sequence $X_0, X_1, \ldots, X_m$ of random variables so that for $0 \leq i < m$,

$$\mathrm{E}(X_{i+1}|X_i, \ldots, X_0) = X_i.$$

Let $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_m$ be a chain of σ-algebras. For $0 \leq i \leq m$, let $X_i = \mathrm{E}(X|\mathcal{F}_i)$. Then $X_0, X_1, \ldots, X_m$ forms a martingale. Typically, $X_0 = \mathrm{E}(X)$ and $X_m = X$.

Martingale

A martingale is a sequence $X_0, X_1, \ldots, X_m$ of random variables so that for $0 \leq i < m$,

$$E(X_{i+1}|X_i, \ldots, X_0) = X_i.$$

Let $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_m$ be a chain of σ-algebras. For $0 \leq i \leq m$, let $X_i = E(X|\mathcal{F}_i)$. Then $X_0, X_1, \ldots, X_m$ forms a martingale. Typically, $X_0 = E(X)$ and $X_m = X$.

- Vertex-exposure Martingale.
- Edge-exposure Martingale.

Azuma's inequality

Theorem: Let $\mathrm{E}(X) = X_0, \ldots, X_m = X$ be a martingale with

$$|X_i - X_{i+1}| \leq 1$$

for all $0 \leq i < m$. For any $\lambda > 0$, Then

$$\Pr(X - \mathrm{E}(X) > \lambda) < e^{-\frac{\lambda^2}{2m}}.$$

Azuma's inequality

Theorem: Let $\mathrm{E}(X) = X_0, \ldots, X_m = X$ be a martingale with

$$|X_i - X_{i+1}| \leq 1$$

for all $0 \leq i < m$. For any $\lambda > 0$, Then

$$\Pr(X - \mathrm{E}(X) > \lambda) < e^{-\frac{\lambda^2}{2m}}.$$

Proof: Let $Y_i = X_i - X_{i-1}$. We have

$$\mathrm{E}(Y_i | X_{i-1}, X_{i-2}, \ldots, X_0) = 0.$$

$$\mathrm{E}(e^{tY_i} | X_{i-1}, X_{i-2}, \ldots, X_0) \leq \cosh(t) \leq e^{t^2/2}.$$

continue

$$
\begin{aligned}
\mathrm{E}(e^{t(X-\mathrm{E}(X))}) &= \mathrm{E}(\prod_{i=1}^{m} e^{tY_i}) \\
&\le \mathrm{E}\left[\left(\prod_{i=1}^{m-1} e^{tY_i}\mathrm{E}(e^{tY_m|X_{m-1},X_{m-2},\ldots,X_0})\right)\right] \\
&\le \mathrm{E}\left[\left(\prod_{i=1}^{m-1} e^{tY_i}\right)\right] e^{t^2/2} \le e^{mt^2/2}.
\end{aligned}
$$

Continue

$$
\begin{aligned}
\Pr(X - \mathrm{E}(X) > \lambda) &= \Pr(e^{t(X-\mathrm{E}(X))} > e^{t\lambda}) \\
&\le e^{-t\lambda}\mathrm{E}(e^{t(X-\mathrm{E}(X))}) \\
&\le e^{-t\lambda + mt^2/2}.
\end{aligned}
$$

Continue

$$\begin{aligned}\Pr(X - \mathrm{E}(X) > \lambda) &= \Pr(e^{t(X-\mathrm{E}(X))} > e^{t\lambda}) \\ &\le e^{-t\lambda}\mathrm{E}(e^{t(X-\mathrm{E}(X))}) \\ &\le e^{-t\lambda+mt^2/2}.\end{aligned}$$

Choose $t = \lambda/m$. We have

$$\Pr(X - \mathrm{E}(X) > \lambda) \le e^{-\frac{\lambda^2}{2m}}.$$

$\square$

Application

- $G = (V, E)$: a simple graph.

Application

- $G = (V, E)$: a simple graph.
- A k-coloring $c\colon V \to [k]$ is **proper** if for any edge uv $c(u) \neq c(v)$.

Application

- $G = (V, E)$: a simple graph.
- A k-coloring $c\colon V \to [k]$ is **proper** if for any edge uv $c(u) \neq c(v)$.
- G is k-**colorable** if there exists a proper k-coloring of G.

Application

- $G = (V, E)$: a simple graph.
- A k-coloring $c\colon V \to [k]$ is **proper** if for any edge uv $c(u) \neq c(v)$.
- G is k-**colorable** if there exists a proper k-coloring of G.
- The **chromatic number** $\chi(G)$ is the minimum integer k such that G.

Application

- $G = (V, E)$: a simple graph.
- A k-coloring $c \colon V \to [k]$ is **proper** if for any edge uv $c(u) \neq c(v)$.
- G is k-**colorable** if there exists a proper k-coloring of G.
- The **chromatic number** $\chi(G)$ is the minimum integer k such that G.

Theorem [Shamir-Spencer (1987)]: For $G = G(n, p)$, we have

$$\Pr(|\chi(G) - \mathrm{E}(\chi(G))| > \lambda\sqrt{n-1}) < 2e^{-\lambda^2/2}.$$

Proof

Let $X = \chi(G)$. Consider the vertex exposure martingale of X: $\mathrm{E}(X) = X_1, \ldots, X_n = X$. Note that for $1 \leq i \leq n$

$$|X_i - X_{i-1}| \leq 1.$$

Proof

Let $X = \chi(G)$. Consider the vertex exposure martingale of X: $\mathrm{E}(X) = X_1, \ldots, X_n = X$. Note that for $1 \leq i \leq n$

$$|X_i - X_{i-1}| \leq 1.$$

Apply Azumar's inequality, we get

$$\Pr(|\chi(G) - \mathrm{E}(\chi(G))| > \lambda\sqrt{n-1}) < 2e^{-\lambda^2/2}.$$

□

Vertex exposure martingale

A graph function f is said to satisfy the **vertex Lipshitz condition** if whenever H and H' differ at only one vertex, $|f(H) - f(H')| \leq 1$. Then

$$\Pr\left(|f(G) - \mathrm{E}(f(G))| > \lambda\sqrt{n-1}\right) < 2e^{-\lambda^2/2}.$$

Vertex exposure martingale

A graph function f is said to satisfy the **vertex Lipshitz condition** if whenever H and H' differ at only one vertex, $|f(H) - f(H')| \leq 1$. Then

$$\Pr\left(|f(G) - \mathrm{E}(f(G))| > \lambda\sqrt{n-1}\right) < 2e^{-\lambda^2/2}.$$

A graph function f is said to satisfy the **edge Lipshitz condition** if whenever H and H' differ at only one edge, $|f(H) - f(H')| \leq 1$. Then

$$\Pr\left(|f(G) - \mathrm{E}(f(G))| > \lambda\sqrt{\binom{n}{2}}\right) < 2e^{-\lambda^2/2}.$$

Tight concentration of $\chi(G)$

For sparse $G = G(n,p)$, there is a better concentration result. Let $p = n^{-\alpha}$.

- **Shamir-Spencer (1987)**: If $\alpha > \frac{5}{6} + \epsilon$, then $\chi(G)$ is concentrated on at most five values.

Tight concentration of $\chi(G)$

For sparse $G = G(n, p)$, there is a better concentration result. Let $p = n^{-\alpha}$.

- **Shamir-Spencer (1987)**: If $\alpha > \frac{5}{6} + \epsilon$, then $\chi(G)$ is concentrated on at most five values.
- **Luczak (1991)**: If $\alpha > \frac{5}{6} + \epsilon$, then $\chi(G)$ is concentrated in at most two values.

Tight concentration of $\chi(G)$

For sparse $G = G(n,p)$, there is a better concentration result. Let $p = n^{-\alpha}$.

- **Shamir-Spencer (1987)**: If $\alpha > \frac{5}{6} + \epsilon$, then $\chi(G)$ is concentrated on at most five values.
- **Luczak (1991)**: If $\alpha > \frac{5}{6} + \epsilon$, then $\chi(G)$ is concentrated in at most two values.
- **Alon-Krivelevich (1997)**: If $\alpha > \frac{1}{2} + \epsilon$, then $\chi(G)$ is concentrated in at most two values.

Tight concentration of $\chi(G)$

For sparse $G = G(n, p)$, there is a better concentration result. Let $p = n^{-\alpha}$.

- **Shamir-Spencer (1987)**: If $\alpha > \frac{5}{6} + \epsilon$, then $\chi(G)$ is concentrated on at most five values.
- **Luczak (1991)**: If $\alpha > \frac{5}{6} + \epsilon$, then $\chi(G)$ is concentrated in at most two values.
- **Alon-Krivelevich (1997)**: If $\alpha > \frac{1}{2} + \epsilon$, then $\chi(G)$ is concentrated in at most two values.

Here we will prove a weaker result.

Theorem: For $\alpha > \frac{5}{6}$ and $p = n^{-\alpha}$, let $G = G(n, p)$. Then $\chi(G)$ is concentrated on at most four values.

A Lemma

Lemma: Let α, c be fixed, $\alpha > \frac{5}{6} + \epsilon$. Let $p = n^{-\alpha}$. Then almost always every $c\sqrt{n}$ vertices of $G = G(n, p)$ may be three-colored.

A Lemma

Lemma: Let α, c be fixed, $\alpha > \frac{5}{6} + \epsilon$. Let $p = n^{-\alpha}$. Then almost always every $c\sqrt{n}$ vertices of $G = G(n, p)$ may be three-colored.

Proof: If not, let T be the minimal set such that is not three-colorable. $G|_T$ has minimum degree at least 3. The probability of existing such T with $|T| < c\sqrt{n}$ is at most

$$\sum_{t=4}^{c\sqrt{n}} \binom{n}{t} \binom{\binom{t}{2}}{3t/2} p^{3t/2} \leq \sum_{t=4}^{c\sqrt{n}} \left(\frac{ne}{t}\right)^t \left(\frac{ne}{3}\right)^{3t/2} p^{3t/2}$$

$$= \sum_{t=4}^{c\sqrt{n}} (c_2 n^{-\epsilon})^t = o(1).$$

Proof of Theorem

Proof: Let $\epsilon > 0$ be arbitrary small and let $u = (n, p, \epsilon)$ be the least integer so that

$$\Pr(\chi(G) \leq u) > \epsilon.$$

Proof of Theorem

Proof: Let $\epsilon > 0$ be arbitrary small and let $u = (n, p, \epsilon)$ be the least integer so that

$$\Pr(\chi(G) \leq u) > \epsilon.$$

Let Y to be the minimal size of a set of vertices S for which $G - S$ may be u-colored. Y satisfies the vertex Lipschitz condition.

Proof of Theorem

Proof: Let $\epsilon > 0$ be arbitrary small and let $u = (n, p, \epsilon)$ be the least integer so that

$$\Pr(\chi(G) \le u) > \epsilon.$$

Let Y to be the minimal size of a set of vertices S for which $G - S$ may be u-colored. Y satisfies the vertex Lipschitz condition. Apply Azuma's inequality with $\lambda = \sqrt{2(n-1)\ln(1/\epsilon)} = O(\sqrt{n})$.

$$\Pr(Y - \mathrm{E}(Y) > \lambda) < \epsilon,$$
$$\Pr(Y - \mathrm{E}(Y) < -\lambda) < \epsilon.$$

Continue

By definition of u, $\Pr(Y = 0) > \epsilon$. Hence $\mathrm{E}(Y) \leq \lambda$.

$$\Pr(Y \geq 2\lambda) \leq \Pr(Y \geq \mathrm{E}(Y) + \lambda) \leq \epsilon.$$

Continue

By definition of u, $\Pr(Y = 0) > \epsilon$. Hence $\mathrm{E}(Y) \leq \lambda$.

$$\Pr(Y \geq 2\lambda) \leq \Pr(Y \geq \mathrm{E}(Y) + \lambda) \leq \epsilon.$$

With probability at least $1 - \epsilon$ there is a u-coloring of all but at most $O(\sqrt{n})$ vertices. By the Lemma, with probability at least $1 - \epsilon$, these points my be colored with three further colors. Thus G is $u + 3$-colorable. Putting together, we have

$$\Pr(u \leq \chi(G) \leq u + 3) \geq 1 - 3\epsilon$$

where ϵ is arbitrarily small. □

Generalization

- $\mathbf{c} := (c_1, \ldots, c_n)$, where $c_i > 0$.
- A martingale $\mathrm{E}(X) = X_0, X_1, \ldots, X_n = X$ is $\mathbf{c}$-Lipschitz if
$$|X_i - X_{i-1}| \leq c_i$$
for $i = 1, 2, \ldots, n$.

Azuma's inequality: If a martingale X is $\mathbf{c}$-Lipschitz, then

$$\Pr(|X - E(X)| \geq \lambda) \leq 2e^{-\frac{\lambda^2}{2\sum_{i=1}^{n} c_i^2}}.$$

Connection

Let $Y_1, Y_2, \ldots, Y_n$ be independent variables and $Y = \sum_{i=1}^{n} Y_i$. Let $X_i = \mathrm{E}(Y) + \sum_{j=1}^{i}(Y_j - \mathrm{E}(Y_j))$. Then $\mathrm{E}(Y) = X_0, X_1, \ldots, X_n = Y$ forms a martingale.

Connection

Let $Y_1, Y_2, \ldots, Y_n$ be independent variables and $Y = \sum_{i=1}^{n} Y_i$. Let $X_i = \mathrm{E}(Y) + \sum_{j=1}^{i}(Y_j - \mathrm{E}(Y_j))$. Then $\mathrm{E}(Y) = X_0, X_1, \ldots, X_n = Y$ forms a martingale.

- Inequalities on martingale can be applied to the sum of independent random variables.

Connection

Let $Y_1, Y_2, \ldots, Y_n$ be independent variables and $Y = \sum_{i=1}^{n} Y_i$. Let $X_i = \mathrm{E}(Y) + \sum_{j=1}^{i}(Y_j - \mathrm{E}(Y_j))$. Then $\mathrm{E}(Y) = X_0, X_1, \ldots, X_n = Y$ forms a martingale.

- Inequalities on martingale can be applied to the sum of independent random variables.
- One may expect to generalize Chernoff-type inequalities to martingales.

Terminologies

We say X is a martingale associated with a filter $\mathbf{F}$ if

- $\mathbf{F} := \{\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_n\}$ is a set of σ-algebras satisfying

$$\{\emptyset, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n.$$

- X is a random variable and it is $\mathcal{F}_n$-measurable.

Terminologies

We say X is a martingale associated with a filter $\mathbf{F}$ if

- $\mathbf{F} := \{\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_n\}$ is a set of σ-algebras satisfying

$$\{\emptyset, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n.$$

- X is a random variable and it is $\mathcal{F}_n$-measurable.

For $1 \leq i \leq n$, let $X_i = \mathrm{E}(X|\mathcal{F}_i)$. Then $X_0, X_1, \ldots, X_n$ forms a martingale.

Terminologies

We say X is a martingale associated with a filter $\mathbf{F}$ if

- $\mathbf{F} := \{\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_n\}$ is a set of σ-algebras satisfying

$$\{\emptyset, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n.$$

- X is a random variable and it is $\mathcal{F}_n$-measurable.

For $1 \leq i \leq n$, let $X_i = \mathrm{E}(X|\mathcal{F}_i)$. Then $X_0, X_1, \ldots, X_n$ forms a martingale.

If a martingale $\mathrm{E}(X_n) = X_0, X_1, \ldots, X_n$ is given, then one can define $\mathcal{F}_i$ be the σ-algebra generated by $X_0, X_1, \ldots, X_i$.

Variation I

Theorem 1: Let X be the martingale associated with a filter $\mathbf{F}$ satisfying

1. $\mathrm{Var}(X_i|\mathcal{F}_{i-1}) \leq \sigma_i^2$, for $1 \leq i \leq n$;
2. $|X_i - X_{i-1}| \leq M$, for $1 \leq i \leq n$.

Then, we have

$$\Pr(X - E(X) \geq \lambda) \leq e^{-\frac{\lambda^2}{2(\sum_{i=1}^n \sigma_i^2 + M\lambda/3)}}.$$

Variation II

Theorem 2: Let X be the martingale associated with a filter $\mathbf{F}$ satisfying

1. $\mathrm{Var}(X_i|\mathcal{F}_{i-1}) \leq \sigma_i^2$, for $1 \leq i \leq n$;
2. $X_i - X_{i-1} \leq M_i$, for $1 \leq i \leq n$.

Then, we have

$$\Pr(X - E(X) \geq \lambda) \leq e^{-\frac{\lambda^2}{2\sum_{i=1}^n (\sigma_i^2 + M_i^2)}}.$$

Variation III

Theorem 3: Let X be the martingale associated with a filter $\mathbf{F}$ satisfying

1. $\mathrm{Var}(X_i|\mathcal{F}_{i-1}) \leq \sigma_i^2$, for $1 \leq i \leq n$;
2. $X_i - X_{i-1} \leq a_i + M$, for $1 \leq i \leq n$.

Then, we have

$$\Pr(X - E(X) \geq \lambda) \leq e^{-\frac{\lambda^2}{2(\sum_{i=1}^n (\sigma_i^2 + a_i^2) + M\lambda/3)}}.$$

Variation IV

Theorem 4: Let X be the martingale associated with a filter $\mathbf{F}$ satisfying

1. $\mathrm{Var}(X_i|\mathcal{F}_{i-1}) \leq \sigma_i^2$, for $1 \leq i \leq n$;
2. $X_i - X_{i-1} \leq M_i$, for $1 \leq i \leq n$.

Then, for any M, we have

$$\Pr(X - E(X) \geq \lambda) \leq e^{-\frac{\lambda^2}{2(\sum_{i=1}^n \sigma_i^2 + \sum_{M_i > M}(M_i - M)^2 + M\lambda/3)}}.$$

The function $g(y)$

$$g(y) = 2\sum_{k=2}^{\infty} \frac{y^{k-2}}{k!} = \frac{2(e^y - 1 - y)}{y^2}.$$

The function $g(y)$

$$g(y) = 2\sum_{k=2}^{\infty} \frac{y^{k-2}}{k!} = \frac{2(e^y - 1 - y)}{y^2}.$$

Facts:

- $g(0) = 1.$

The function $g(y)$

$$g(y) = 2\sum_{k=2}^{\infty} \frac{y^{k-2}}{k!} = \frac{2(e^y - 1 - y)}{y^2}.$$

Facts:

- $g(0) = 1.$
- $g(y) \leq 1$, for $y < 0$.

The function $g(y)$

$$g(y) = 2\sum_{k=2}^{\infty} \frac{y^{k-2}}{k!} = \frac{2(e^y - 1 - y)}{y^2}.$$

Facts:

- $g(0) = 1.$
- $g(y) \leq 1$, for $y < 0.$
- $g(y)$ is monotone increasing, for $y \geq 0.$

The function $g(y)$

$$g(y) = 2\sum_{k=2}^{\infty} \frac{y^{k-2}}{k!} = \frac{2(e^y - 1 - y)}{y^2}.$$

Facts:

- $g(0) = 1$.
- $g(y) \leq 1$, for $y < 0$.
- $g(y)$ is monotone increasing, for $y \geq 0$.
- For $y < 3$, we have

$$g(y) = 2\sum_{k=2}^{\infty} \frac{y^{k-2}}{k!} \leq \sum_{k=2}^{\infty} \frac{y^{k-2}}{3^{k-2}} = \frac{1}{1 - y/3}.$$

Proof of Theorem 3

Since $\mathrm{E}(X_i|\mathcal{F}_{i-1}) = X_{i-1}$ and $X_i - X_{i-1} - a_i \leq M$, we have

$$
\begin{aligned}
&\mathrm{E}(e^{t(X_i - X_{i-1} - a_i)}|\mathcal{F}_{i-1}) \\
&= \mathrm{E}(\sum_{k=0}^{\infty} \frac{t^k}{k!}(X_i - X_{i-1} - a_i)^k|\mathcal{F}_{i-1}) \\
&= 1 - ta_i + \mathrm{E}(\sum_{k=2}^{\infty} \frac{t^k}{k!}(X_i - X_{i-1} - a_i)^k|\mathcal{F}_{i-1}) \\
&\leq 1 - ta_i + \mathrm{E}(\frac{t^2}{2}(X_i - X_{i-1} - a_i)^2 g(tM)|\mathcal{F}_{i-1}) \\
&= 1 - ta_i + \frac{t^2}{2} g(tM)\mathrm{E}((X_i - X_{i-1} - a_i)^2|\mathcal{F}_{i-1})
\end{aligned}
$$

Continue

$$
\begin{aligned}
\mathrm{E}(e^{t(X_i - X_{i-1} - a_i)} | \mathcal{F}_{i-1}) \\
&\le 1 - ta_i + \frac{t^2}{2} g(tM) \mathrm{E}((X_i - X_{i-1} - a_i)^2 | \mathcal{F}_{i-1}) \\
&= 1 - ta_i + \frac{t^2}{2} g(tM) (\mathrm{E}((X_i - X_{i-1})^2 | \mathcal{F}_{i-1}) + a_i^2) \\
&\le 1 - ta_i + \frac{t^2}{2} g(tM) (\sigma_i^2 + a_i^2) \\
&\le e^{-ta_i + \frac{t^2}{2} g(tM)(\sigma_i^2 + a_i^2)}.
\end{aligned}
$$

Continue

$$
\begin{aligned}
&\mathrm{E}(e^{t(X_i - X_{i-1} - a_i)} | \mathcal{F}_{i-1}) \\
&\quad \leq \; 1 - ta_i + \frac{t^2}{2} g(tM) \mathrm{E}((X_i - X_{i-1} - a_i)^2 | \mathcal{F}_{i-1}) \\
&\quad = \; 1 - ta_i + \frac{t^2}{2} g(tM) (\mathrm{E}((X_i - X_{i-1})^2 | \mathcal{F}_{i-1}) + a_i^2) \\
&\quad \leq \; 1 - ta_i + \frac{t^2}{2} g(tM) (\sigma_i^2 + a_i^2) \\
&\quad \leq \; e^{-ta_i + \frac{t^2}{2} g(tM)(\sigma_i^2 + a_i^2)}.
\end{aligned}
$$

Thus, $\mathrm{E}(e^{tX_i} | \mathcal{F}_{i-1})$

$$
\begin{aligned}
\mathrm{E}(e^{tX_i} | \mathcal{F}_{i-1}) &= \mathrm{E}(e^{t(X_i - X_{i-1} - a_i)} | \mathcal{F}_{i-1}) e^{tX_{i-1} + ta_i} \\
&\leq e^{-ta_i + \frac{t^2}{2} g(tM)(\sigma_i^2 + a_i^2)} e^{tX_{i-1} + ta_i} \\
&= e^{\frac{t^2}{2} g(tM)(\sigma_i^2 + a_i^2)} e^{tX_{i-1}}.
\end{aligned}
$$

Continue

Inductively, we have

$$
\begin{aligned}
\mathrm{E}(e^{tX}) &= \mathrm{E}(\mathrm{E}(e^{tX_n}|\mathcal{F}_{n-1})) \\
&\le e^{\frac{t^2}{2}g(tM)(\sigma_n^2+a_n^2)}\mathrm{E}(e^{tX_{n-1}}) \\
&\le \cdots \\
&\le \prod_{i=1}^{n} e^{\frac{t^2}{2}g(tM)(\sigma_i^2+a_i^2)}\mathrm{E}(e^{tX_0}) \\
&= e^{\frac{1}{2}t^2 g(tM)\sum_{i=1}^{n}(\sigma_i^2+a_i^2)}e^{t\mathrm{E}(X)}.
\end{aligned}
$$

Continue

Then for t satisfying $tM < 3$, we have

$$\begin{aligned}\Pr(X \ge \mathrm{E}(X) + \lambda) &= \Pr(e^{tX} \ge e^{t\mathrm{E}(X)+t\lambda}) \\ &\le e^{-t\mathrm{E}(X)-t\lambda}\mathrm{E}(e^{tX}) \\ &\le e^{-t\lambda}e^{\frac{1}{2}t^2 g(tM)\sum_{i=1}^n(\sigma_i^2+a_i^2)} \\ &= e^{-t\lambda+\frac{1}{2}t^2 g(tM)\sum_{i=1}^n(\sigma_i^2+a_i^2)} \\ &\le e^{-t\lambda+\frac{1}{2}\frac{t^2}{1-tM/3}\sum_{i=1}^n(\sigma_i^2+a_i^2)}.\end{aligned}$$

We choose $t = \frac{\lambda}{\sum_{i=1}^n(\sigma_i^2+a_i^2)+M\lambda/3}$. Clearly $tM < 3$ and

$$\begin{aligned}\Pr(X \ge \mathrm{E}(X) + \lambda) &\le e^{-t\lambda+\frac{1}{2}\frac{t^2}{1-tM/3}\sum_{i=1}^n(\sigma_i^2+a_i^2)} \\ &= e^{-\frac{\lambda^2}{2(\sum_{i=1}^n(\sigma_i^2+a_i^2)+M\lambda/3)}}. \qquad \square\end{aligned}$$

More variations

- Generalized to sub-martingales and super-martingales.
- Generalized to a sequence of random variables which is close to martingales.

Reference

- C. McDiarmid. Concentration. In Probabilistic Methods for Algorithmic Discrete Mathematics, edited by M. Habib, C. McDiarmid, J. Ramirez- Alfonsin, and B. Reed, pp. 195248, Algorithms and Combinatorics 16. Berlin: Springer, 1998.
- Chung and Lu, Concentration inequalities and martingale inequalities — a survey, Internet Mathematics, 3 (2006), No. 1, 79-127.
- Chung and Lu, Complex Graphs and Networks, (2006) published by AMS, ISBN-10: 0-8218-3657-9, ISBN-13: 978-0-8218-3657-6.