Probabilistic Methods in Combinatorics
Lecture 10

Linyuan Lu
University of South Carolina

Mathematical Sciences Center at Tsinghua University
November 16, 2011 – December 30, 2011
A martingale is a sequence $X_0, X_1, \ldots, X_m$ of random variables so that for $0 \leq i < m$,

$$E(X_{i+1} | X_i, \ldots, X_0) = X_i.$$
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 $\sigma$-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 $X_m = X$. 
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 $\sigma$-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.**
**Theorem:** Let $E(X) = X_0, \ldots, X_m = X$ be a martingale with

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

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

$$\Pr(X - E(X) > \lambda) < e^{-\frac{\lambda^2}{2m}}.$$
Azuma’s inequality

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

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

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

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

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

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

$$E(e^{tY_i}|X_{i-1}, X_{i-2}, \ldots, X_0) \leq \cosh(t) \leq e^{t^2/2}.$$
\[
E(e^{t(X - E(X))}) = E \left( \prod_{i=1}^{m} e^{tY_i} \right) \\
\leq E \left[ \left( \prod_{i=1}^{m-1} e^{tY_i} E(e^{tY_m | X_{m-1}, X_{m-2}, \ldots, X_{0}}) \right) \right] \\
\leq E \left[ \left( \prod_{i=1}^{m-1} e^{tY_i} \right) e^{t^2/2} \right] e^{mt^2/2} \leq e^{mt^2/2}.
\]
\[ \Pr(X - E(X) > \lambda) = \Pr(e^{t(X - E(X))} > e^{t\lambda}) \leq e^{-t\lambda}E(e^{t(X - E(X))}) \leq e^{-t\lambda + mt^2/2}. \]
\[ \Pr(X - E(X) > \lambda) = \Pr(e^{t(X-E(X))} > e^{t\lambda}) \leq e^{-t\lambda}E(e^{t(X-E(X))}) \leq e^{-t\lambda + mt^2/2}. \]

Choose \( t = \lambda/m \). We have

\[ \Pr(X - E(X) > \lambda) \leq e^{-\frac{\lambda^2}{2m}}. \]
Application

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

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

- $G = (V, E)$: a simple graph.
- A $k$-coloring $c: V \rightarrow [k]$ 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 $c : V \rightarrow [k]$ 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 $c: V \rightarrow [k]$ 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

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

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

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

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

Apply Azumar’s inequality, we get

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

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

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

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

$$\Pr \left( |f(G) - E(f(G))| > \lambda \sqrt{\binom{n}{2}} \right) < 2e^{-\lambda^2/2}.$$
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.
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.
Lemma: Let $\alpha, c$ 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 $\alpha, c$ 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

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

$$
= \sum_{t=4}^{c\sqrt{n}} \left( c2n^{-\epsilon} \right)^{t} = o(1).
$$
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: 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: 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. Apply Azuma’s inequality with

$$\lambda = \sqrt{2(n - 1) \ln(1/\epsilon)} = O(\sqrt{n}).$$

Then,

$$\Pr(Y - E(Y) > \lambda) < \epsilon,$$

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

\[
\Pr(Y \geq 2\lambda) \leq \Pr(Y \geq \mathbb{E}(Y) + \lambda) \leq \epsilon.
\]
By definition of $u$, $\Pr(Y = 0) > \epsilon$. Hence $E(Y) \leq \lambda$.

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

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 may 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 $\epsilon$ is arbitrarily small.
Generalization

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

Azuma’s inequality: If a martingale \( X \) is \( c \)-Lipschitz, then

\[
\Pr(|X - E(X)| \geq \lambda) \leq 2e^{-\frac{\lambda^2}{2\sum_{i=1}^{n} c_i^2}}.
\]
Let $Y_1, Y_2, \ldots, Y_n$ be independent variables and $Y = \sum_{i=1}^{n} Y_i$. Let $X_i = E(Y) + \sum_{j=1}^{i} (Y_j - E(Y_j))$. Then $E(Y) = X_0, X_1, \ldots, X_n = Y$ forms a martingale.
Let $Y_1, Y_2, \ldots, Y_n$ be independent variables and 
$Y = \sum_{i=1}^{n} Y_i$. Let $X_i = E(Y) + \sum_{j=1}^{i}(Y_j - E(Y_j))$. Then 
$E(Y) = X_0, X_1, \ldots, X_n = Y$ forms a martingale.

- Inequalities on martingale can be applied to the sum of 
independent random variables.
Let $Y_1, Y_2, \ldots, Y_n$ be independent variables and $Y = \sum_{i=1}^{n} Y_i$. Let $X_i = E(Y) + \sum_{j=1}^{i} (Y_j - E(Y_j))$. Then $E(Y) = X_0, X_1, \ldots, X_n = Y$ 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 $\mathcal{F}$ if

- $\mathcal{F} := \{\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_n\}$ is a set of $\sigma$-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.
We say $X$ is a martingale associated with a filter $\mathcal{F}$ if

- $\mathcal{F} := \{\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_n\}$ is a set of $\sigma$-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 $X_i = \mathbb{E}(X|\mathcal{F}_i)$. Then $X_0, X_1, \ldots, X_n$ forms a martingale.
We say $X$ is a martingale associated with a filter $F$ if

- $F := \{\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_n\}$ is a set of $\sigma$-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 $X_i = E(X|\mathcal{F}_i)$. Then $X_0, X_1, \ldots, X_n$ forms a martingale.

If a martingale $E(X_n) = X_0, X_1, \ldots, X_n$ is given, then one can define $\mathcal{F}_i$ be the $\sigma$-algebra generated by $X_0, X_1, \ldots, X_i$. 
**Theorem 1:** Let $X$ be the martingale associated with a filter $F$ satisfying

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

Then, we have

$$\Pr(X - E(X) \geq \lambda) \leq e^{-\frac{\lambda^2}{2(\sum_{i=1}^{n} \sigma_i^2 + M\lambda/3)}}.$$
**Theorem 2:** Let $X$ be the martingale associated with a filter $F$ satisfying

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

Then, we have

$$\Pr(X - E(X) \geq \lambda) \leq e^{-\frac{\lambda^2}{2 \sum_{i=1}^{n}(\sigma_i^2 + M_i^2)}}.$$
**Theorem 3:** Let $X$ be the martingale associated with a filter $F$ satisfying

1. $\text{Var}(X_i|\mathcal{F}_{i-1}) \leq \sigma_i^2$, 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

$$\Pr(X - E(X) \geq \lambda) \leq e^{-\frac{\lambda^2}{2 \left( \sum_{i=1}^{n} (\sigma_i^2 + a_i^2) + M\lambda/3 \right)}}.$$
**Theorem 4:** Let $X$ be the martingale associated with a filter $\mathcal{F}$ satisfying

1. $\text{Var}(X_i|\mathcal{F}_{i-1}) \leq \sigma_i^2$, 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 $E(X_i|\mathcal{F}_{i-1}) = X_{i-1}$ and $X_i - X_{i-1} - a_i \leq M$, we have

$$E(e^{t(X_i-X_{i-1}-a_i)}|\mathcal{F}_{i-1}) = E(\sum_{k=0}^{\infty} \frac{t^k}{k!} (X_i - X_{i-1} - a_i)^k|\mathcal{F}_{i-1})$$

$$= 1 - ta_i + E(\sum_{k=2}^{\infty} \frac{t^k}{k!} (X_i - X_{i-1} - a_i)^k|\mathcal{F}_{i-1})$$

$$\leq 1 - ta_i + E(\frac{t^2}{2} (X_i - X_{i-1} - a_i)^2 g(tM)|\mathcal{F}_{i-1})$$

$$= 1 - ta_i + \frac{t^2}{2} g(tM) E(((X_i - X_{i-1} - a_i)^2|\mathcal{F}_{i-1})$$
\[
\begin{align*}
E(e^{t(X_i - X_{i-1} - a_i)}|\mathcal{F}_{i-1}) & \leq 1 - ta_i + \frac{t^2}{2} g(tM) E((X_i - X_{i-1} - a_i)^2|\mathcal{F}_{i-1}) \\
& = 1 - ta_i + \frac{t^2}{2} g(tM) (E((X_i - X_{i-1})^2|\mathcal{F}_{i-1}) + a_i^2) \\
& \leq 1 - ta_i + \frac{t^2}{2} g(tM) (\sigma_i^2 + a_i^2) \\
& \leq e^{-ta_i + \frac{t^2}{2} g(tM)(\sigma_i^2 + a_i^2)}.
\end{align*}
\]
\[ \mathbb{E}(e^{t(X_i - X_{i-1} - a_i)}|\mathcal{F}_{i-1}) \]
\[ \leq 1 - ta_i + \frac{t^2}{2} g(tM) \mathbb{E}((X_i - X_{i-1} - a_i)^2|\mathcal{F}_{i-1}) \]
\[ = 1 - ta_i + \frac{t^2}{2} g(tM) \mathbb{E}((X_i - X_{i-1})^2|\mathcal{F}_{i-1}) + a_i^2 \]
\[ \leq 1 - ta_i + \frac{t^2}{2} g(tM) (\sigma_i^2 + a_i^2) \]
\[ \leq e^{-ta_i + \frac{t^2}{2} g(tM) (\sigma_i^2 + a_i^2)}. \]

Thus, \[ \mathbb{E}(e^{tX_i}|\mathcal{F}_{i-1}) = \mathbb{E}(e^{t(X_i - X_{i-1} - a_i)}|\mathcal{F}_{i-1}) e^{tX_{i-1} + ta_i} \]
\[ \leq e^{-ta_i + \frac{t^2}{2} g(tM) (\sigma_i^2 + a_i^2)} e^{tX_{i-1} + ta_i} \]
\[ = e^{\frac{t^2}{2} g(tM) (\sigma_i^2 + a_i^2)} e^{tX_{i-1}}. \]
Inductively, we have

\[
E(e^{tX}) = E \left( E \left( e^{tX_n} | \mathcal{F}_{n-1} \right) \right) \\
\leq e^{\frac{t^2}{2} g(tM) (\sigma_n^2 + a_n^2)} E(e^{tX_{n-1}}) \\
\leq \cdots \\
\leq \prod_{i=1}^{n} e^{\frac{t^2}{2} g(tM) (\sigma_i^2 + a_i^2)} E(e^{tX_0}) \\
= e^{\frac{1}{2} t^2 g(tM) \sum_{i=1}^{n} (\sigma_i^2 + a_i^2)} e^{tE(X)}. 
\]
Then for \( t \) satisfying \( tM < 3 \), we have

\[
\Pr(X \geq \mathbb{E}(X) + \lambda) = \Pr(e^{tX} \geq e^{t\mathbb{E}(X)+t\lambda})
\]

\[
\leq e^{-t\mathbb{E}(X)-t\lambda} \mathbb{E}(e^{tX})
\]

\[
\leq e^{-t\lambda} e^{t^2g(tM) \sum_{i=1}^{n}(\sigma_i^2+a_i^2)}
\]

\[
= e^{-t\lambda+\frac{1}{2}t^2g(tM) \sum_{i=1}^{n}(\sigma_i^2+a_i^2)}
\]

\[
\leq e^{-t\lambda+\frac{1}{2} \frac{t^2}{1-tM/3} \sum_{i=1}^{n}(\sigma_i^2+a_i^2)}. 
\]

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

\[
\Pr(X \geq \mathbb{E}(X) + \lambda) \leq e^{-t\lambda+\frac{1}{2} \frac{t^2}{1-tM/3} \sum_{i=1}^{n}(\sigma_i^2+a_i^2)}
\]

\[
= e^{-\frac{\lambda^2}{2(\sum_{i=1}^{n}(\sigma_i^2+a_i^2)+M\lambda/3)}}. \quad \square
\]
More variations

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