Martingales

Piotr Wojciechowski[1]

[1]Lane Department of Computer Science and Electrical Engineering
West Virginia University

28 April, 2012

Outline

Outline

Outline

We can apply a Chernoff-like bound to Martingale even when the variables are not independent. This bound is referred to as the Azuma-Hoeffding Inequality.

We can apply a Chernoff-like bound to Martingale even when the variables are not independent. This bound is referred to as the Azuma-Hoeffding Inequality.

Azuma-Hoeffding Inequality Theorem

If $X_0, \ldots, X_n$ is a martingale such that $|X_k - X_{k-1}| \leq c_k$ then, for all $t > 0$ and $\lambda > 0$

$$P(|X_t - X_0| \geq \lambda) \leq 2 \cdot e^{-\frac{\lambda^2}{2 \cdot \sum_{k=1}^{t} c_k^2}}.$$

Proof.

Let $Y_i = X_i - X_{i-1}$ for $i = 1, \ldots, t$. Thus $|Y_i| \leq c_i$.

Proof.

Let $Y_i = X_i - X_{i-1}$ for $i = 1, \ldots, t$. Thus $|Y_i| \leq c_i$. Since $X_0, \ldots, X_n$ is a martingale we have that

$$E[Y_i \,|\, X_0, \ldots, X_{i-1}] = E[X_i - X_{i-1} \,|\, X_0, \ldots, X_{i-1}] = E[X_i \,|\, X_0, \ldots, X_{i-1}] - X_{i-1} = 0.$$

Proof.

Let $Y_i = X_i - X_{i-1}$ for $i = 1, \ldots, t$. Thus $|Y_i| \leq c_i$. Since $X_0, \ldots, X_n$ is a martingale we have that

$$E[Y_i \,|\, X_0, \ldots, X_{i-1}] = E[X_i - X_{i-1} \,|\, X_0, \ldots, X_{i-1}] = E[X_i \,|\, X_0, \ldots, X_{i-1}] - X_{i-1} = 0.$$

We have that

$$Y_i = -c_i \cdot \frac{1 - \frac{Y_i}{c_i}}{2} + c_i \cdot \frac{1 + \frac{Y_i}{c_i}}{2}.$$

Proof.

Let $Y_i = X_i - X_{i-1}$ for $i = 1, \ldots, t$. Thus $|Y_i| \le c_i$. Since $X_0, \ldots, X_n$ is a martingale we have that

$$E[Y_i \mid X_0, \ldots, X_{i-1}] = E[X_i - X_{i-1} \mid X_0, \ldots, X_{i-1}] = E[X_i \mid X_0, \ldots, X_{i-1}] - X_{i-1} = 0.$$

We have that

$$Y_i = -c_i \cdot \frac{1 - \frac{Y_i}{c_i}}{2} + c_i \cdot \frac{1 + \frac{Y_i}{c_i}}{2}.$$

Because the function $e^{\alpha \cdot x}$ is convex it follows that

$$e^{\alpha \cdot Y_i} \le \frac{1 - \frac{Y_i}{c_i}}{2} \cdot e^{-\alpha \cdot c_i} + \frac{1 + \frac{Y_i}{c_i}}{2} \cdot e^{\alpha \cdot c_i} = \frac{e^{\alpha \cdot c_i} + e^{-\alpha \cdot c_i}}{2} + \frac{Y_i}{2 \cdot c_i} \cdot (e^{\alpha \cdot c_i} - e^{-\alpha \cdot c_i}).$$

□

Proof.

As shown previously $E[Y_i \mid X_0, \ldots, X_{i-1}] = 0$, thus

$$\begin{aligned} E[e^{\alpha \cdot Y_i} \mid X_0, \ldots, x_{i-1}] &= E\left[\frac{e^{\alpha \cdot c_i} + e^{-\alpha \cdot c_i}}{2} + \frac{Y_i}{2 \cdot c_i} \cdot (e^{\alpha \cdot c_i} - e^{-\alpha \cdot c_i}) \mid X_0, .., X_{i-1}\right] \\ &= \frac{e^{\alpha \cdot c_i} + e^{-\alpha \cdot c_i}}{2} \leq e^{-\frac{(\alpha \cdot c_i)^2}{2}}. \end{aligned}$$

Proof.

As shown previously $E[Y_i \mid X_0, \ldots, X_{i-1}] = 0$, thus

$$E[e^{\alpha \cdot Y_i} \mid X_0, \ldots, x_{i-1}] = E\left[\frac{e^{\alpha \cdot c_i} + e^{-\alpha \cdot c_i}}{2} + \frac{Y_i}{2 \cdot c_i} \cdot (e^{\alpha \cdot c_i} - e^{-\alpha \cdot c_i}) \mid X_0, .., X_{i-1}\right]$$
$$= \frac{e^{\alpha \cdot c_i} + e^{-\alpha \cdot c_i}}{2} \leq e^{-\frac{(\alpha \cdot c_i)^2}{2}}.$$

We have that

$$E\left[e^{\alpha \cdot (X_t - X_0)}\right] = E\left[\prod_{i=1}^{t-1} e^{\alpha \cdot Y_i}\right] = E\left[\prod_{i=1}^{t-2} e^{\alpha \cdot Y_i}\right] \cdot E\left[e^{\alpha \cdot Y_{i-1}} \mid X_0, \ldots, x_{i-1}\right]$$
$$\leq E\left[\prod_{i=1}^{t-2} e^{\alpha \cdot Y_i}\right] \cdot e^{\frac{(\alpha \cdot c_i)^2}{2}} \leq e^{\frac{\alpha^2 \cdot \sum_{k=1}^{t} c_i^2}{2}}.$$

□

Proof.

Therefore, we have that

$$P(X_t - X_0 \geq \lambda) = P(e^{\alpha\cdot(X_t - X_0)} \geq e^{\alpha\cdot\lambda}) \leq \frac{E\left[e^{\alpha\cdot(X_t - X_0)}\right]}{e^{\alpha\cdot\lambda}} \leq e^{\frac{\alpha^2\cdot\sum_{k=1}^{t} c_i^2}{2} - \alpha\cdot\lambda}.$$

Proof.

Therefore, we have that

$$P(X_t - X_0 \geq \lambda) = P(e^{\alpha \cdot (X_t - X_0)} \geq e^{\alpha \cdot \lambda}) \leq \frac{E\left[e^{\alpha \cdot (X_t - X_0)}\right]}{e^{\alpha \cdot \lambda}} \leq e^{\frac{\alpha^2 \cdot \sum_{k=1}^{t} c_i^2}{2} - \alpha \cdot \lambda}.$$

By letting $\alpha = \frac{\lambda}{\sum_{k=1}^{t} c_i^2}$ we get that

$$P(X_t - X_0 \geq \lambda) \leq e^{-\frac{\lambda^2}{2 \cdot \sum_{k=1}^{t} c_k^2}}.$$

Proof.

Therefore, we have that

$$P(X_t - X_0 \geq \lambda) = P(e^{\alpha\cdot(X_t - X_0)} \geq e^{\alpha\cdot\lambda}) \leq \frac{E\left[e^{\alpha\cdot(X_t - X_0)}\right]}{e^{\alpha\cdot\lambda}} \leq e^{\frac{\alpha^2\cdot\sum_{k=1}^{t} c_i^2}{2} - \alpha\cdot\lambda}.$$

By letting $\alpha = \frac{\lambda}{\sum_{k=1}^{t} c_i^2}$ we get that

$$P(X_t - X_0 \geq \lambda) \leq e^{-\frac{\lambda^2}{2\cdot\sum_{k=1}^{t} c_k^2}}.$$

We can similarly construct the same bound on $P(X_t - X_0 \leq -\lambda)$. □

Corollary

If $X_0, \ldots, X_n$ is a martingale such that $|X_k - X_{k-1}| \leq c$ then, for all $t \geq 1$ and $\lambda > 0$

$$P(|X_t - X_0| \geq \lambda \cdot c \cdot \sqrt{t}) \leq 2 \cdot e^{-\frac{\lambda^2}{2}}.$$

Corollary

If $X_0, \ldots, X_n$ is a martingale such that $|X_k - X_{k-1}| \leq c$ then, for all $t \geq 1$ and $\lambda > 0$

$$P(|X_t - X_0| \geq \lambda \cdot c \cdot \sqrt{t}) \leq 2 \cdot e^{-\frac{\lambda^2}{2}}.$$

Azuma-Hoeffding Inequality Theorem

If $X_0, \ldots, X_n$ is a martingale such that $B_k \leq X_k - X_{k-1} \leq B_k + d_k$ for some constants d_k and random variables B_k then, for all $t \geq 0$ and $\lambda > 0$

$$P(|X_t - X_0| \geq \lambda) \leq 2 \cdot e^{-\frac{2 \cdot \lambda^2}{\sum_{k=1}^{t} d_k^2}}.$$

Outline

Lipschitz condition

A function $f(\bar{X}) = f(X_1, X_2, \ldots, X_n)$ satisfies the *Lipschitz condition* with bound c if for any i and any $x_1, \ldots, x_n$ and y_i,

$$|f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n)| \leq c.$$

Lipschitz condition

A function $f(\bar{X}) = f(X_1, X_2, \ldots, X_n)$ satisfies the *Lipschitz condition* with bound c if for any i and any $x_1, \ldots, x_n$ and y_i,

$$|f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n)| \leq c.$$

Theorem

Let f be a functions satisfying the Lipschitz condition with bound c and let $Z_0, \ldots$ be the Doob martingale defined by $Z_0 = E[f(X_1, \ldots, X_n)]$ and $Z_k = E[f(X_1, \ldots, X_n) \,|\, X_1, \ldots, X_k]$. We have that for each k there exists a random variable B_k depending on $Z_0, \ldots, Z_{k-1}$ such that $B_k \leq Z_k - Z_{k-1} \leq B_k + c$.

Balls and Bins

Balls and Bins Example

Suppose that we are throwing m balls independently and uniformly at random into n bins. Let X_i denote the bin into which the i^{th} ball falls and let F denote the number of empty bins after all m balls are thrown.

Balls and Bins

Balls and Bins Example

Suppose that we are throwing m balls independently and uniformly at random into n bins. Let X_i denote the bin into which the i^{th} ball falls and let F denote the number of empty bins after all m balls are thrown. We will use the Azuma-Hoeffding Inequality to determine the tightness of the distribution of F.

Balls and Bins

Balls and Bins Example

Suppose that we are throwing m balls independently and uniformly at random into n bins. Let X_i denote the bin into which the i^{th} ball falls and let F denote the number of empty bins after all m balls are thrown. We will use the Azuma-Hoeffding Inequality to determine the tightness of the distribution of F.

We have that the sequence $Z_i = E[F \mid X_1, \ldots, X_i]$ is Doob martingale.

Balls and Bins

Balls and Bins Example

Suppose that we are throwing m balls independently and uniformly at random into n bins. Let X_i denote the bin into which the i^{th} ball falls and let F denote the number of empty bins after all m balls are thrown. We will use the Azuma-Hoeffding Inequality to determine the tightness of the distribution of F.

We have that the sequence $Z_i = E[F \mid X_1, \ldots, X_i]$ is Doob martingale.

Since F depends on $X_1, \ldots, X_m$, there is a function f such that $F = f(X_1, \ldots, X_m)$.

Balls and Bins

Balls and Bins Example

Suppose that we are throwing m balls independently and uniformly at random into n bins. Let X_i denote the bin into which the i^{th} ball falls and let F denote the number of empty bins after all m balls are thrown. We will use the Azuma-Hoeffding Inequality to determine the tightness of the distribution of F.
We have that the sequence $Z_i = E[F \mid X_1, \ldots, X_i]$ is Doob martingale.
Since F depends on $X_1, \ldots, X_m$, there is a function f such that $F = f(X_1, \ldots, X_m)$.
Because changing which bin a single ball lands in changes F by at most 1, we have that f satisfies the Lipschitz condition with bound 1.

Balls and Bins

Balls and Bins Example

Suppose that we are throwing m balls independently and uniformly at random into n bins. Let X_i denote the bin into which the i^{th} ball falls and let F denote the number of empty bins after all m balls are thrown. We will use the Azuma-Hoeffding Inequality to determine the tightness of the distribution of F.

We have that the sequence $Z_i = E[F \mid X_1, \ldots, X_i]$ is Doob martingale.

Since F depends on $X_1, \ldots, X_m$, there is a function f such that $F = f(X_1, \ldots, X_m)$.

Because changing which bin a single ball lands in changes F by at most 1, we have that f satisfies the Lipschitz condition with bound 1.

Thus applying the second Azuma-Hoeffding Inequality we get that

$$P(|F - E[F]| \geq \epsilon) = P(|Z_m - Z_0| \geq \epsilon) \leq 2 \cdot e^{-\frac{2 \cdot \epsilon^2}{\sum_{k=1}^{m} c^2}} = e^{-\frac{2 \cdot \epsilon^2}{m}}.$$

Chromatic Number

Chromatic Number Example

Let G be a random graph in $G_{n,p'}$. The *Chromatic number*, $\chi(G)$, is the minimum number of colors needed to color all the verticies of a graph so that no two adjacent verticies are the same color.

Chromatic Number

Chromatic Number Example

Let G be a random graph in $G_{n,p'}$. The *Chromatic number*, $\chi(G)$, is the minimum number of colors needed to color all the verticies of a graph so that no two adjacent verticies are the same color. We will use the Azuma-Hoeffding Inequality to determine the tightness of the distribution of $\chi(G)$.

Chromatic Number

Chromatic Number Example

Let G be a random graph in $G_{n,p'}$. The *Chromatic number*, $\chi(G)$, is the minimum number of colors needed to color all the verticies of a graph so that no two adjacent verticies are the same color. We will use the Azuma-Hoeffding Inequality to determine the tightness of the distribution of $\chi(G)$.

Let Z be the vertex exposure martingale for G. Thus if G_i is subgraph of G induced by the verticies $1, \ldots, i$, we have that $Z_0 = E[\chi(G)]$ and $Z_i = E[\chi(G) \mid G_1, \ldots, G_i]$.

Chromatic Number

Chromatic Number Example

Let G be a random graph in $G_{n,p'}$. The *Chromatic number*, $\chi(G)$, is the minimum number of colors needed to color all the verticies of a graph so that no two adjacent verticies are the same color. We will use the Azuma-Hoeffding Inequality to determine the tightness of the distribution of $\chi(G)$.
Let Z be the vertex exposure martingale for G. Thus if G_i is subgraph of G induced by the verticies $1, \ldots, i$, we have that $Z_0 = E[\chi(G)]$ and $Z_i = E[\chi(G) \mid G_1, \ldots, G_i]$.
Introducing a vertex into the graph increases he chromatic number by at most 1 so we have that for each i, $0 \leq Z_i - Z_{i-1} \leq 1$.

Chromatic Number

Chromatic Number Example

Let G be a random graph in $G_{n,p'}$. The *Chromatic number*, $\chi(G)$, is the minimum number of colors needed to color all the verticies of a graph so that no two adjacent verticies are the same color. We will use the Azuma-Hoeffding Inequality to determine the tightness of the distribution of $\chi(G)$.

Let Z be the vertex exposure martingale for G. Thus if G_i is subgraph of G induced by the verticies $1, \ldots, i$, we have that $Z_0 = E[\chi(G)]$ and $Z_i = E[\chi(G) \mid G_1, \ldots, G_i]$. Introducing a vertex into the graph increases he chromatic number by at most 1 so we have that for each i, $0 \leq Z_i - Z_{i-1} \leq 1$. Thus, we can apply the second Azuma-Hoeffding Inequality to get that

$$P(|\chi(G) - E[\chi(G)]| \geq \lambda \cdot \sqrt{n}) \leq 2 \cdot e^{-2 \cdot \lambda^2}.$$