Martingales

Martingales are useful in handling sums of random variables which are not totally independent.

They are useful in deriving tight concentration (Chernoff-like) bounds on the values of functions of dependent random variables.

Conditional Expectation

Definition 1. *The random variable* $E[X|Y]$ *is defined to be the random variable* $f(Y)$ *such that* $f(y) = E[X|Y = y]$ *which is the conditional expectation of* X *given that* $Y = y$*:*

$$E[X|Y = y] = \frac{\sum_x xp(x, y)}{\sum_x p(x, y)}$$

Example: Consider independent throws of an unbiased 6-sided die. For $1 \leq i \leq 6$, let X_i denote the number of times the value i appears in n throws of the die. Then

$$E[X_1|X_2] = \frac{n - X_2}{5}$$

$$E[X_1|X_2, X_3] = \frac{n - X_2 - X_3}{4}$$

Properties of Conditional Expectation:

1. $E[E[X|Y]] = E[X]$
2. $E[E[X|Y, Z]|Y] = E[X|Y]$

Martingale Definition

Definition 2. *A sequence of random variables* $Z_0, Z_1, \ldots$ *is a* martingale with respect to the sequence $X_0, X_1, \ldots$, *if for all* $n \geq 0$, *the following conditions hold:*

1. Z_n *is a function of* $X_0, \ldots, X_n$*;*

2. $E[|Z_n|] < \infty$.

3. $E[Z_{n+1}|X_0, \ldots, X_n] = Z_n$.

A sequence of random variables $Z_0, Z_1, \ldots,$ *is said to be a* martingale *if it is a martingale with respect to itself. That is,* $E[|Z_n|] < \infty$, *and, for all* $n \geq 0$, $E[Z_{n+1}|Z_0, \ldots, Z_n] = Z_n$.

Example

Consider a gambler who plays a sequence of fair games. Let X_i be the amount the gambler wins on the ith game, and Z_i be the gambler's total winnings at the end of the ith game. Since $E[X_i] = 0$,

$$E[Z_{i+1}|X_1, X_2, \ldots, X_i] = Z_i + E[X_{i+1}] = Z_i.$$

Thus, $Z_1, Z_2, \ldots$ is a martingale w.r.t the sequence $X_1, X_2, \ldots$.

Doob Martingale

Let $X_0, \ldots, X_n$ be *any* sequence of random variables and let Y be any random variable with $E[|Y|] < \infty$.

Define the random variable $Z_i = E[Y | X_0, \ldots, X_i]$, $i = 0, 1, \ldots, n$.

Then $Z_0, Z_1, \ldots, Z_n$ form a martingale sequence w.r.t. $X_0, X_1, \ldots X_n$ called as a **Doob martingale**.

$$E[Z_{i+1} | X_0, \ldots, X_i] = E[E[Y | X_0, \ldots, X_{i+1}] | X_0, \ldots, X_i]$$

$$= E[Y | X_0, \ldots, X_i]$$
$$= Z_i$$

Examples of Doob Martingale

Example 1: Let G be a random graph from $G(n,p)$. Label the $m = \binom{n}{2}$ possible edge slots in some arbitrary order, and let the 0-1 r.v. X_j be the indicator for an edge to be in the jth slot.

Let F be any finite-valued function defined over graphs. Let

$Z_0 = E[F(G)]$ and

$Z_i = E[F(G)|X_1, \ldots, X_i]$, $i = 1, \ldots, m$.

$Z_0, Z_1, \ldots Z_m$ is a Doob martingale called the **edge exposure martingale**.

Example 2: Let G be a random graph from $G(n, p)$. Fix an arbitrary numbering of the vertices from 1 through n, and let G_i be the subgraph of G induced by the first i vertices.

Let $Z_0 = E[F(G)]$ and

$Z_i = E[F(G)|G_1, \ldots, G_i]$, $i = 1, \ldots, n$.

$Z_0, Z_1, \ldots Z_m$ is a Doob martingale called the **vertex exposure martingale**.

Azuma-Hoeffding Inequality

Let $X_0, X_1, \ldots$ be a martingale sequence such that for each k,

$$|X_k - X_{k-1}| \leq c_k$$

where c_k may depend on k. Then for all $t \geq 0$ and any $\lambda > 0$,

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

Corollary 1. *Let $X_0, X_1, \ldots$ be a martingale sequence such that, for all $k \geq 1$, If $|X_k - X_{k-1}| \leq c$ where c is independent of k, then for all $t \geq 1$ and any $\lambda > 0$,*

$$\Pr(|X_t - X_0| \geq \lambda c\sqrt{t}) \leq 2e^{-\lambda^2/2}$$

"Method of Bounded Differences".

Application

Suppose we have a sequence of random variables $X_1, \ldots, X_n$ and a function $f(X_1, \ldots, X_n)$.

Define the Doob martingale: $Y_0, Y_1, \ldots, Y_n$ where, $Y_0 = E[f(X_1, \ldots, X_n)]$, and,

$Y_i = E[f(X_1, \ldots, X_n) | X_1, \ldots, X_i]$, $1 \leq i \leq n$.

Suppose $|Y_i - Y_{i-1}| \leq c$, for all $1 \leq i \leq n$.

Then we use Azuma's inequality to show a concentration bound for $f(X_1, \ldots, X_n)$.

Example – A Chernoff-like bound

Let $Z_1, \ldots, Z_n$ be 0-1, independent, random variables that take value 1 with probability p.

We obtain a tight concentration bound $S = \sum_{i=1}^{n} Z_i$.

Define a martingale sequence $X_0, X_1, \ldots, X_n$ where:

$X_0 = E[S]$ and $X_i = E[S|Z_1, \ldots, Z_i]$, $i = 1, \ldots, n$.

$E[X_i|X_0, \ldots, X_{i-1}] = E[E[S|Z_1, \ldots, Z_i]|X_0, \ldots, X_{i-1}]$

$= E[S|X_0, \ldots, X_{i-1}]$

$= E[S|Z_1, \ldots, Z_{i-1}] = X_{i-1}$.

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

By Azuma's inequality,

$\Pr(|X - E[X]| \geq \lambda) = \Pr(|X_n - X_0| \geq \lambda) \leq 2e^{-\frac{\lambda^2}{2n}}$.

Example: Chromatic number

Theorem 1. *Let $X(G)$ be the chromatic number of a random graph G in $G(n,p)$. Then, for any $\lambda > 0$,*

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

Proof: Consider the vertex-exposure martingale:

Let $Z_0 = E[X(G)]$ and

$Z_i = E[X(G)|G_1, \ldots, G_i]$, $i = 1, \ldots, n$.

$|Z_i - Z_{i-1}| \leq 1$.

Proof of Azuma-Hoeffding Inequality

We upper bound $E[e^{\alpha(X_t - X_0)}]$.

Define $Y_i = X_i - X_{i-1}$, $i = 1, \ldots, t$.

Then, $X_t - X_0 = \sum_{i=1}^{t} Y_i$.

$|Y_i| \leq c$, and since $X_0, X_1, \ldots$ is a martingale,

$$E[Y_i | X_0, X_1, \ldots, X_{i-1}] = E[X_i - X_{i-1} | X_0, X_1, \ldots, X_{i-1}]$$

$$= E[X_i | X_0, X_1, \ldots, X_{i-1}] - X_{i-1} = 0.$$

We consider $E[e^{\alpha Y_i} | X_0, X_1, \ldots, X_{i-1}]$.

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

Since $e^{\alpha Y_i}$ is convex for all $\alpha > 0$,

$e^{\alpha Y_i} \leq e^{-\alpha c_i} \frac{1 - Y_i/c_i}{2} + e^{\alpha c_i} \frac{1+Y_i}{2}$

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

Hence,

$$E[e^{\alpha Y_i}|X_0, X_1, \ldots, X_{i-1}]$$

$$\leq E[\frac{e^{\alpha c_i} + e^{-\alpha c_i}}{2} + \frac{Y_i}{2c_i}(e^{\alpha c_i} - e^{-\alpha c_i})|X_0, X_1, \ldots, X_{i-1}]$$
$$= \frac{e^{\alpha c_i} + e^{-\alpha c_i}}{2}$$
$$\leq e^{(\alpha c_i)^2/2}.$$

$$E[e^{\alpha(X_t - X_0)}] = E[\Pi_{i=1}^{t} e^{\alpha Y_i}]$$

$$E[E[\Pi_{i=1}^{t} e^{\alpha Y_i}|X_0, X_1, \ldots, X_{t-1}]]$$

$$= E[\Pi_{i=1}^{t-1} e^{\alpha Y_i} E[e^{\alpha Y_t}|X_0, X_1, \ldots, X_{t-1}]]$$

$$\leq E[\Pi_{i=1}^{t-1} e^{\alpha Y_i}] e^{(\alpha c_t)^2/2}$$
$$\leq e^{\alpha^2 \sum_{k=1}^{t} c_k^2/2}.$$

Thus,

$$\Pr(X_t - X_0 \geq \lambda) = \Pr(e^{\alpha(X_t - X_0)} \geq e^{\alpha\lambda})$$

$$\leq \frac{E[e^{\alpha(X_t - X_0)}]}{e^{\alpha\lambda}}$$

$$\leq e^{\alpha^2 \sum_{k=1}^{t} c_k^2/2 - \alpha\lambda}$$

$$\leq e^{-\frac{\lambda^2}{2\sum_{k=1}^{t} c_k^2}}$$

by choosing $\alpha = \frac{\lambda}{\sum_{k=1}^{t} c_k^2}$.

A similar argument gives the bound for $\Pr(X_t - X_0 \leq -\lambda)$.

Lipschitz condition

Let $f : D_1 \times \ldots \times D_n \rightarrow R$ be a real-valued function with n arguments from possibly distinct domains.

The function f is said to satisfy the **Lipschitz condition** with bound c, if for any $x_1 \in D_1, \ldots, x_n \in D_n$, any $i \in \{1, \ldots, n\}$, and any $y_i \in D_i$,

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

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

That is, changing the value of any single coordinate can change the function value by at most c.

A General Framework for Applying Azuma's Inequality

Theorem 2. *Suppose we have a sequence of random variables* $X_1, \ldots, X_n$ *and a function* $f(\hat{X}) = f(X_1, \ldots, X_n)$ *which satisfies Lipschitz condition.*

Consider the Doob martingale:

$Z_0 = E[f(X_1, \ldots, X_n)]$ *and*

$Z_k = E[f(X_1, \ldots, X_n) | X_1, \ldots, X_k]$, $1 \leq k \leq n$.

If $X_1, \ldots, X_n$ *are independent random variables, then* $|Z_k - Z_{k-1}| \leq c$.

We can thus use Azuma to get a concentration bound for f.

Proof

We will show for discrete random variables that only take a finite number of values.

Let S_k stand for $X_1, X_2, \ldots, X_k$.

And $f_k(\hat{X}, x) = f(X_1, \ldots, X_{k-1}, x, X_{k+1}, \ldots, X_n)$.

and $f_k(\hat{z}, x) = f(z_1, \ldots, z_{k-1}, x, z_{k+1}, \ldots, z_n)$.

$Z_k - Z_{k-1} = E[f(\hat{X})|S_k] - E[f(\hat{X})|S_{k-1}]$

$Z_k - Z_{k-1} \leq max_x E[f(\hat{X}|S_{k-1}, X_k = x] - E[f(\hat{X})|S_{k-1}]$

$Z_k - Z_{k-1} \geq min_y E[f(\hat{X})|S_{k-1}, X_k = y] - E[f(\hat{X})|S_{k-1}]$

We bound:

$max_x E[f(\hat{X})|S_{k-1}, X_k = x] - min_y E[f(\hat{X}|S_{k-1}, X_k = y]$

$= max_{x,y}(E[f(\hat{X})|S_{k-1}, X_k = x] - E[f(\hat{X}|S_{k-1}, X_k = y])$

$= max_{x,y}(E[f_k(\hat{X}, x) - f_k(\hat{X}, y)|S_{k-1}])$

For any values $x, y, z_1, \ldots, z_{k-1}$,

$E[f_k(\hat{X}, x) - f_k(\hat{X}, y)|X_1 = z_1, \ldots, X_{k-1} = z_{k-1}]$

$= \sum_{z_{k+1},\ldots,z_n} \Pr((X_{k+1} = z_{k+1}) \cap \ldots \cap (X_n = z_n)).(f_k(\hat{z}, x) - f_k(\hat{z}, y)).$

$\leq c$.

Hence $E[f_k(\hat{X}, x) - f_k(\hat{X}, y)|S_{k-1}] \leq c$.

Balls and Bins

Suppose m balls are thrown independently and uniformly into n bins.

Let Z be the number of empty bins.

$\mu = E[Z] = n(1 - \frac{1}{n})^m \sim ne^{m/n}$.

We want to show a concentration result for Z.

Let X_t represent the bin into which the tth ball falls.

The sequence $Z_t = E[Z|X_1, \ldots, X_t]$ is a Doob martigale.

$Z = F(X_1, \ldots, X_n)$ satisfies the Lipschitz condition.

Hence, $|Z_t - Z_{t-1}| \leq 1$.

Thus, $\Pr(|Z - E[Z]| \geq \lambda) \leq 2e^{-\lambda^2/2m}$.

A Stronger Concentration

Theorem 3. *Let $r = m/n$. Then for any $\lambda > 0$,*

$$\Pr(|Z - \mu| \geq \lambda) \leq 2e^{-\frac{\lambda^2(n-1/2)}{n^2-\mu^2}}.$$

Proof:

We bound $|Z_t - Z_{t-1}|$ more carefully.

Define $z(Y,t)$ as the expectation of Z given that Y bins are empty at time t.

$z(Y,t) = E[Z|Y \text{ bins are empty at time } t]$

$= Y(1-1/n)^{m-t}$.

Let Y_t denote the number of empty bins at time t.

$Z_{t-1} = z(Y_{t-1}, t-1) = Y_{t-1}(1-1/n)^{m-t+1}$.

At time t there are two possibilities:

1. The tth ball goes into a currently non-empty bin.

Then $Y_t = Y_{t-1}$, and

$Z_t = z(Y_t, t) = z(Y_{t-1}, t) = Y_{t-1}(1 - 1/n)^{m-t}$.

2. The t ball goes into a currently empty bin.

Then $Y_t = Y_{t-1} - 1$ and

$Z_t = z(Y_t, t) = z(Y_{t-1} - 1, t) = (Y_{t-1} - 1)(1 - 1/n)^{m-t}$.

We bound $\Delta_t = Z_t - Z_{t-1}$.

1. With probability $1 - Y_{t-1}/n$, Δ_t is

$Y_{t-1}(1 - 1/n)^{m-t} - Y_{t-1}(1 - 1/n)^{m-t+1}$

$= \frac{Y_{t-1}}{n}(1 - 1/n)^{m-t}$

2. With probability Y_{t-1}/n, Δ_t is

$(Y_{t-1} - 1)(1 - 1/n)^{m-t} - Y_{t-1}(1 - 1/n)^{m-t+1}$

$= -(1 - \frac{Y_{t-1}}{n})(1 - 1/n)^{m-t}$

Thus,

$-(1 - 1/n)^{m-t} \leq \Delta_t \leq (1 - 1/n)^{m-t}$.

Thus $|Z_t - Z_{t-1}| \leq c_t$ where $c_t = (1 - 1/n)^{m-t}$.

$\sum_{t=1}^{m} c_t^2 = \frac{n^2 - \mu^2}{2n-1}$

Azuma gives the result.