Hoeffding's Bound

Theorem

Let $X_1, \ldots, X_n$ *be independent random variables with* $\mathbf{E}[X_i] = \mu_i$ *and* $Pr(B_i \leq X_i \leq B_i + c_i) = 1$, *then*

$$Pr\left(\left|\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i\right| \geq \epsilon\right) \leq e^{-\frac{2\epsilon^2}{\sum_{i=1}^{n} c_i^2}}$$

Do we need independence?

Martingales

Definition

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 hold:

1. Z_n is a function of $X_0, X_1, \ldots, X_n$;
2. $\mathbf{E}[|Z_n|] < \infty$;
3. $\mathbf{E}[Z_{n+1} | X_0, X_1, \ldots, X_n] = Z_n$;

Definition

A sequence of random variables $Z_0, Z_1, \ldots$ is a *martingale* when it is a martingale with respect to itself, that is

1. $\mathbf{E}[|Z_n|] < \infty$;
2. $\mathbf{E}[Z_{n+1} | Z_0, Z_1, \ldots, Z_n] = Z_n$;

Example

A series of fair games ($\mathbf{E}[gain] = 0$), not necessarily independent..

Game 1: bet $\$1$.

Game $i > 1$: bet 2^i if won in round $i - 1$; bet i otherwise.

$X_i =$ amount won in ith game. ($X_i < 0$ if ith game lost).

$Z_i =$ total winnings at end of ith game.

Example

$X_i =$ amount won in ith game. ($X_i < 0$ if ith game lost).

$Z_i =$ total winnings at end of ith game.

$Z_1, Z_2, \ldots$ is martingale with respect to $X_1, X_2, \ldots$

$\mathbf{E}[X_i] = 0.$

$\mathbf{E}[Z_i] = \sum \mathbf{E}[X_i] = 0 < \infty.$

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

Doob Martingale

Let $X_0, X_1, \dots, X_n$ be sequence of random variables. Let $Y = f(X_1, \dots, X_n)$ be a random variable with $\mathbf{E}[|Y|] < \infty$.

Let $Z_0 = \mathbf{E}[Y]$

Let $Z_i = \mathbf{E}[Y | X_0, X_1, \dots, X_i]$, $i = 0, 1, \dots, n$

$Z_0, Z_1, \dots, Z_n$ is martingale with respect to $X_0, X_1, \dots, X_n$.

Proof

Lemma

$\mathbf{E}[\mathbf{E}[V|U,W]|W] = \mathbf{E}[V|W]$.

$Z_i = \mathbf{E}[Y|X_0, X_1, \ldots, X_i]$, $i = 0, 1, \ldots, n$

$$\begin{aligned} \mathbf{E}[Z_{i+1}|X_0, X_1, \ldots, X_i] &= \mathbf{E}[\mathbf{E}[Y|X_0, X_1, \ldots, X_{i+1}]|X_0, X_1, \ldots, X_i] \\ &= \mathbf{E}[Y|X_0, X_1, \ldots, X_i] \\ &= Z_i. \end{aligned}$$

Example: Edge Exposure Martingale

Let G random graph from $G_{n,p}$. Consider $m = \binom{n}{2}$ possible edges in arbitrary order.

$$X_i = \begin{cases} 1 & \text{if } i\text{th edge is present} \\ 0 & \text{otherwise} \end{cases}$$

$F(G) =$ size maximum clique in G.

$Z_0 = \mathbf{E}[F(G)]$

$Z_i = \mathbf{E}[F(G)|X_1, X_2, \ldots, X_i]$, for $i = 1, \ldots, m$.

$Z_0, Z_1, \ldots, Z_m$ is a Doob martingale.

($F(G)$ could be any finite-valued function on graphs.)

Tail Inequalities

Theorem (Azuma-Hoeffding Inequality)

Let $Z_0, Z_1, \ldots, Z_n$ be a martingale (with respect to $X_1, X_2, \ldots$) such that $|Z_k - Z_{k-1}| \leq c_k$. Then, for all $t \geq 0$ and any $\lambda > 0$

$$\Pr(|Z_t - Z_0| \geq \lambda) \leq 2e^{-\lambda^2/(2\sum_{k=1}^{t} c_k^2)}.$$

The following corollary is often easier to apply.

Corollary

Let $X_0, X_1, \ldots$ be a martingale such that for all $k \geq 1$,

$$|X_k - X_{k-1}| \leq c.$$

Then for all $t \geq 1$ and $\lambda > 0$,

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

Tail Inequalities: A More General Form

Theorem (Azuma-Hoeffding Inequality)

Let $Z_0, Z_1, \ldots, Z_n$ be a martingale with respect to $X_1, X_2 \ldots$, such that

$$B_k \leq Z_k - Z_{k-1} \leq B_k + c_k$$

for some constants c_k and for some random variables B_k that may be functions of $X_0, X_1, \ldots, X_{k-1}$. Then, for all $t \geq 0$ and any $\lambda > 0$

$$\Pr(|Z_t - Z_0| \geq \lambda) \leq 2e^{-2\lambda^2/(\sum_{k=1}^{t} c_k^2)}.$$

Proof

Let $X^k = X_1, \ldots, X_k$ and $Z_{i+1} - Z_i = X_i$,
By Hoeffding's Lemma: $E[e^{\lambda X_i} \mid X^{i-1}] \leq e^{\lambda^2 c_i^2/8}$.

$$\begin{aligned}
\mathbf{E}[e^{\lambda \sum_{i=1}^n X_i}] &= \mathbf{E}\left[[\mathbf{E}[e^{\lambda \sum_{i=1}^n X_i} \mid X^{n-1}]\right] \\
&= \mathbf{E}\left[e^{\lambda \sum_{i=1}^{n-1} X_i} \mathbf{E}[e^{\lambda X_n} \mid X^{n-1}]\right] \\
&\leq e^{\lambda^2 c_n^2/8} \mathbf{E}\left[e^{\lambda \sum_{i=1}^{n-1} X_i}\right] \\
&\leq e^{\lambda^2 \sum_{i=1}^n c_i^2/8}
\end{aligned}$$

$$Pr(|Z_t - Z_0| \geq \lambda) \leq e^{-\lambda\epsilon} e^{\lambda^2 \sum_{i=1}^n c_i^2/8} \leq 2e^{-2\epsilon^2/(\sum_{k=1}^t c_k^2)},$$

For $\lambda = \frac{4\epsilon}{\sum_{i=1}^n c_i^2}$.

Tail Inequalities: Doob Martingales

Let $X_1, \ldots, X_n$ be sequence of random variables.

Random variable Y:

- Y is a function of $X_1, X_2, \ldots, X_n$;
- $\mathbf{E}[|Y|] < \infty$.

Let $Z_i = \mathbf{E}[Y | X_1, \ldots, X_i]$, $i = 0, 1, \ldots, n$.

$Z_0, Z_1, \ldots, Z_n$ is martingale with respect to $X_1, \ldots, X_n$.

If we can use Azuma-Hoeffding inequality:

$$\Pr(|Z_n - Z_0| \geq \lambda) \leq .....$$

then we have,

$$\Pr(|Y - \mathbf{E}[Y]| \geq \lambda) \leq ......$$

Example: Pattern Matching

Given a string and a pattern: is the pattern interesting?

Does it appear more often than is expected in a random string?

Is the number of occurrences of the pattern concentrated around the expectation?

$A = (a_1, a_2, \ldots, a_n)$ string of characters, each chosen independently and uniformly at random from Σ, with $s = |\Sigma|$.

pattern: $B = (b_1, \ldots, b_k)$ fixed string, $b_i \in \Sigma$.

F= number occurrences of B in random string A.

$$\mathbf{E}[F] = ?$$

$A = (a_1, a_2, \ldots, a_n)$ string of characters, each chosen independently and uniformly at random from Σ, with $m = |\Sigma|$.

pattern: $B = (b_1, \ldots, b_k)$ fixed string, $b_i \in \Sigma$.

$F=$ number occurrences of B in random string S.

$$\mathbf{E}[F] = (n - k + 1)\left(\frac{1}{m}\right)^k .$$

Can we bound the deviation of F from its expectation?

$F=$ number occurrences of B in random string A.

$Z_0 = \mathbf{E}[F]$

$Z_i = \mathbf{E}[F|a_1, \ldots, a_i]$, for $i = 1, \ldots, n$.

$Z_0, Z_1, \ldots, Z_n$ is a Doob martingale.

$Z_n = F$.

$F=$ number occurrences of B in random string A.

$Z_0 = \mathbf{E}[F]$

$Z_i = \mathbf{E}[F | a_1, \ldots, a_i]$, for $i = 1, \ldots, n$.

$Z_0, Z_1, \ldots, Z_n$ is a Doob martingale.

$Z_n = F$.

Each character in A can participate in no more than k occurrences of B:

$$|Z_i - Z_{i+1}| \le k.$$

Azuma-Hoeffding inequality (version 1):

$$\Pr(|F - \mathbf{E}[F]| \ge \lambda) \le 2e^{-\lambda^2/(2nk^2)}.$$

McDiarmid Bound

$f(X_1, X_2, \ldots, X_n)$ satisfies *Lipschitz condition* with bound c if for any i and any set of values $x_1, \ldots, x_n$ and y:

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

Then,

$$\begin{aligned} \Pr(|Z_n - Z_0| \geq \lambda) &= \Pr(|f(\ldots) - \mathbf{E}[f(\ldots)]| \geq \lambda) \\ &\leq 2e^{-2\lambda^2/(\sum_{k=1}^{n} c_k^2)}. \end{aligned}$$

Lemma

If $X_1, \ldots, X_n$ are independent then $|Z_i - Z_{i-1}| \leq c_i$.

$$Z_k - Z_{k-1} \quad = \quad \mathbf{E}[f(\bar{X}) \mid X^k] - \mathbf{E}[f(\bar{X}) \mid X^{k-1}].$$

Hence $Z_k - Z_{k-1}$ is bounded above by

$$\sup_x \mathbf{E}[f(\bar{X}) \mid X^{k-1}, X_k = x] - \mathbf{E}[f(\bar{X}) \mid X^{k-1}]$$

and bounded below by

$$\inf_y \mathbf{E}[f(\bar{X}) \mid X^{k-1}, X_k = y] - \mathbf{E}[f(\bar{X}) \mid X^{k-1}].$$

Therefore, if we let

$$B_k = \inf_y \mathbf{E}[f(\bar{X}) \mid X^{k-1}, X_k = y] - \mathbf{E}[f(\bar{X}) \mid X^{k-1}],$$

then if we can bound

$$\sup_x \mathbf{E}[f(\bar{X}) \mid X^{k-1}, X_k = x] - \inf_y \mathbf{E}[f(\bar{X}) \mid X^{k-1}, X_k = y] \leq c,$$

then we will have appropriately bounded the gap $Z_k - Z_{k-1}$.

$$\begin{aligned} Z_k - Z_{k-1} &= \sup_{x,y} \Big(\mathbf{E}[f(\bar{X}) \mid X^{k-1}, X_k = x] - \mathbf{E}[f(\bar{X}) \mid X^{k-1}, X_k = y] \Big) \\ &= \sup_{x,y} \mathbf{E}[f(\bar{X}, x) - f(\bar{X}, y) \mid X^{k-1}]. \end{aligned}$$

Because the X_i are independent, the probability of any specific set of values for X_{k+1} through X_n does not depend on the values of $X_1, \ldots, X_k$. Hence, for any values $z_1, \ldots, z_{k-1}$ we have that

$$\sup_{x,y} \mathbf{E}[f(\bar{X}, x) - f(\bar{X}, y) \mid X_1 = z_1, \ldots, X_{k-1} = z_{k-1}]$$

is equal to

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

But

$$f(\bar{z}, x) - f(\bar{z}, y) \leq c,$$

and hence so is

$$\mathbf{E}[f(\bar{X}, x) - f(\bar{X}, y) \mid X^{k-1}],$$

giving the required bound.

Application: Balls and Bins

We are throwing m balls independently and uniformly at random into n bins.
Let $X_i =$ the bin that the ith ball falls into.
Let F be the number of empty bins after the m balls are thrown.
Then the sequence

$$Z_i = \mathbf{E}[F \mid X_1, \ldots, X_i]$$

is a Doob martingale.
$F = f(X_1, X_2, \ldots, X_n)$ satisfies the Lipschitz condition with bound 1, thus $|Z_{i+1} - Z_i| \leq 1$
We therefore obtain

$$\Pr(|F - \mathbf{E}[F]| \geq \epsilon) \leq 2\mathrm{e}^{-\epsilon^2/2m}$$

Here

$$\mathbf{E}[F] = n\left(1 - \frac{1}{n}\right)^m,$$

but we could obtain the concentration result without knowing $\mathbf{E}[F]$.

Application: Chromatic Number

Given a random graph G in $G_{n,p}$, the *chromatic number* $\chi(G)$ is the minimum number of colors needed in order to color all vertices of the graph so that no adjacent vertices have the same color.
We use the vertex exposure martingale defined
Let G_i be the random subgraph of G induced by the set of vertices $1, \ldots, i$, let $Z_0 = \mathbf{E}[\chi(G)]$, and let

$$Z_i = \mathbf{E}[\chi(G) \mid G_1, \ldots, G_i].$$

Since a vertex uses no more than one new color, again we have that the gap between Z_i and Z_{i-1} is at most 1,
We conclude

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

This result holds even without knowing $\mathbf{E}[\chi(G)]$.

Example: Edge Exposure Martingale

Let G random graph from $G_{n,p}$. Consider $m = \binom{n}{2}$ possible edges in arbitrary order.

$$X_i = \begin{cases} 1 & \text{if } i\text{th edge is present} \\ 0 & \text{otherwise} \end{cases}$$

$F(G) =$ size maximum clique in G.

$Z_0 = \mathbf{E}[F(G)]$

$Z_i = \mathbf{E}[F(G)|X_1, X_2, \ldots, X_i]$, for $i = 1, \ldots, m$.

$Z_0, Z_1, \ldots, Z_m$ is a Doob martingale.

($F(G)$ could be any finite-valued function on graphs.)