Concentration from Martingales

A sequence of random variables $X_0, X_1, \ldots X_n, \ldots$

where $X_i = X_i(A_0, A_1, \ldots, A_i)$

is called a martingale w.r.t. $A_0, A_1, \ldots, A_n, \ldots$

Nb: $\underbrace{E(X_{i+1} \mid A_0, A_1, \ldots, A_i)}_{\text{this is a random variable.}} = X_i$

$$X_{i+1}(\omega) = \sum_{\substack{\hat{\omega} \,:\, A_j(\hat{\omega}) = A_j(\omega) \\ 0 \le j \le i}} X_{i+1}(\hat{\omega}) \Pr(\hat{\omega}).$$

Theorem

Suppose that $X_0, X_1, \ldots, X_n$ is a martingale, w.r.t. $A_0, A_1, \ldots, A_n$ and

$$a_i \leq X_{i+1} - X_i \leq b_i \quad , \quad i = 1, 2, \ldots, n.$$

Then

$$P_r(|X_n - X_0| \geq t) \leq 2e^{-2t^2/\sum_i (b_i - a_i)^2}.$$

Proof

We first consider $P_r(X_n - X_0 \geq t)$

Fix $\lambda > 0$. Then

$$\Pr(X_n - \mu \geq t) = \Pr\left(e^{\lambda(X_n - X_0 - t)} \geq 1 \right)$$

$$\leq E\left(e^{\lambda(X_n - X_0 - t)} \right)$$

$$= e^{-\lambda t} E\left(e^{\lambda(X_n - X_0)} \right)$$

$$= e^{-\lambda t} E\left(\exp\left\{ \sum_{i=0}^{n} \lambda Y_i \right\} \right)$$

where $Y_i = X_i - X_{i-1}$.

$$= e^{-\lambda t} E\left(\prod_{i=0}^{n} e^{\lambda Y_i} \right)$$

We show that

$$E\left(\prod_{i=0}^{n} e^{\lambda Y_i}\right) \leq e^{\frac{\lambda^2}{8}(b_n-a_n)^2} E\left(\prod_{i=0}^{n-1} e^{\lambda Y_i}\right) \quad (1)$$

and then induction gives

$$E\left(\prod_{i=0}^{n} e^{\lambda Y_i}\right) \leq \exp\left\{\lambda^2 \sum_{i=1}^{n} (b_i-a_i)^2/8\right\}$$

and

$$Pr(X_n - X_0 \geq t) \leq \exp\left\{-\lambda t + \lambda^2 \sum_{i=1}^{n} (b_i-a_i)^2/8\right\}$$

Now choose

$$\lambda = \frac{4t}{\sum_{i=1}^{n} (b_i-a_i)^2}.$$

Proof of (1).

$$E\left(\prod_{i=0}^{n} e^{\lambda Y_i}\right)$$

$$= E\left(\prod_{i=0}^{n-1} e^{\lambda Y_i}\, E\left(e^{\lambda Y_n} \mid A_0, A_1, \dots, A_{n-1}\right)\right)$$

and we obtain (1) from

$$E\left(e^{\lambda Y_n} \mid A_0, A_1, \dots, A_{n-1}\right) \leq e^{\lambda^2 (b_n - a_n)^2 / 8} \qquad (2)$$

Proof of (2)

Y_n satisfies

(i) $E(Y_n) = 0$ and (ii) $a_n \le Y_n \le b_n$

($E(Y_n) = \underset{A_0,\dots,A_{n-1}}{E}\left(\underset{A_n}{E}\left(X_n - X_{n-1} \mid A_0, A_1, \dots, A_{n-1}\right)\right) = 0$)

Now if $a_n \le Y_n \le b_n$

$$e^{\lambda Y_n} \le \frac{Y_n - a_n}{b_n - a_n} e^{\lambda b_n} + \frac{b_n - Y_n}{b_n - a_n} e^{\lambda a_n}$$

and so

$$E(e^{\lambda Y_n}) \le \frac{b_n}{b_n - a_n} e^{\lambda a_n} - \frac{a_n}{b_n - a_n} e^{\lambda b_n}$$

$$E(e^{\lambda Y_n}) \leqslant \frac{b_n}{b_n - a_n} e^{\lambda a_n} - \frac{a_n}{b_n - a_n} e^{\lambda b_n}$$

$$= e^{f(y)}$$

where if $p = -a_n/(b_n - a_n)$, $y = (b_n - a_n)\lambda$,

$$f(y) = -py + \ln(1 - p + pe^{y})$$

$$f''(y) = \frac{p(1-p)e^{-y}}{(p + (1-p)e^{-y})^2} \leqslant \frac{1}{4} \qquad \left[\frac{AB}{(A+B)^2} \leqslant \frac{1}{4}\right]$$

and so

$$f(y) \leqslant y^2/8$$

proving (2).

For the lower bound

$$\Pr(X_n - X_0 \leq -t) = \Pr(-X_n + X_0 \geq t)$$

$$\leq e^{-2t^2 / \sum_i (b_i - a_i)^2} .$$

$\square$

Sometimes our sequence is a submartingale or a supermartingale.

$$E(X_{i+1} | \cdots) \leq X_i \qquad E(X_{i+1} | \cdots) \geq X_i$$

To bound $\Pr(X_n - X_0 \geq t)$ we used

$$e^{\lambda Y_n} \leq \frac{Y_n - a_n}{b_n - a_n} e^{\lambda b_n} + \frac{b_n - Y_n}{b_n - a_n} e^{\lambda a_n}$$

$$= Y_n \Bigg[\underbrace{\frac{e^{\lambda b_n} - e^{\lambda a_n}}{b_n - a_n}}_{\geq 0} \Bigg] + \frac{b_n}{b_n - a_n} e^{\lambda a_n} - \frac{a_n}{b_n - a_n} e^{\lambda b_n}$$

$E(Y_n) \leq 0$ for a supermartingale.

So our estimate for $\Pr(X_n - X_0 \geq t)$ is valid for supermartingales. For $\Pr(X_n - X_0 \leq -t)$, it is valid for submartingales.

We now prove a similar, but slightly different version:

<u>Theorem</u>

Suppose that $X_0, X_1, \dots, X_n$ is a martingale, w.r.t. $A_0, A_1, \dots, A_n$ and for $0 \le a_i \le 1$, $i = 1, 2, \dots, n$,

$$-a_i \le X_{i+1} - X_i \le 1 - a_i \quad , \; i = 1, 2, \dots, n.$$

Let $a = \frac{1}{n}(a_1 + \dots + a_n)$ and $\bar{a} = 1 - a$. Then

$$\Pr(|X_n - X_0| \geq nt) \le \left(\left(\frac{a}{a+t}\right)^{a+t} \left(\frac{\bar{a}}{\bar{a}-t}\right)^{\bar{a}-t} \right)^n$$

for $t \le \bar{a}$.

We will first observe that

$$E(e^{\lambda Y_n}) \leq (1-a_n)e^{-\lambda a_n} + a_n e^{\lambda(1-a_n)}$$

so that

$$P_x(X_n - X_0 \geq nt) \leq e^{-\lambda nt} \prod_{k=1}^{n} \left[(1-a_k)e^{-\lambda a_k} + a_k e^{\lambda(1-a_k)} \right]$$

$$= e^{-\lambda n(a+t)} \prod_{k=1}^{n} (1 - a_k + a_k e^{\lambda})$$

$$\leq e^{-\lambda n(a+t)} (1 - a + a e^{\lambda})^n$$

Now put

$$e^{\lambda} = \frac{(a+t)(1-a)}{a(1-a-t)} .$$

Corollary

Under the conditions above:

(i)

$$\Pr(|X_n - X_0| \geq t) \leq 2e^{-2t^2/n}$$

(ii) $$\Pr(X_n - X_0 \geq \epsilon a n) \leq \left((1+\epsilon)^{1+\epsilon} e^{-\epsilon}\right)^{an}$$

$$\leq e^{-\frac{\epsilon^2 a n}{2(1+\epsilon/3)}}$$

(iii) $$\Pr(X_n - X_0 \leq -\epsilon a n) \leq e^{-\epsilon^2 a n/2}$$

.

(1)

Let

$$f(t) = \ln\left(\left(\frac{a}{a+t}\right)^{a+t}\left(\frac{\bar{a}}{\bar{a}-t}\right)^{\bar{a}-t}\right)$$

$$f'(t) = \ln\left(\frac{a(\bar{a}-t)}{(a+t)\,\bar{a}}\right)$$

$$f''(t) = -\left((a+t)(\bar{a}-t)\right)^{-1} \leqslant -4$$

$f(0) = f'(0) = 0$ and so

$$f(t) \leq -2t^2.$$

(ii)

$$\Pr(X_n - X_0 \geq \epsilon a n) \leq \left[e^{-\lambda a(1+\epsilon)}(1-a+ae^{\lambda})\right]^n$$

Now let $e^{\lambda} = 1+\epsilon$

$$\leq \left[(1+\epsilon)^{-a(1+\epsilon)}(1+a\epsilon)\right]^n$$

$$\leq \left[(1+\epsilon)^{-(1+\epsilon)}e^{\epsilon}\right]^{an}$$

and now use

$$(1+\epsilon)\ln(1+\epsilon) - \epsilon \geq \frac{3\epsilon^2}{6+2\epsilon}$$

to get second inequality in (ii).

(iii)

$$f(t) = \ln\left(\left(\frac{a}{a+t}\right)^{a+t}\left(\frac{\bar{a}}{\bar{a}-t}\right)^{\bar{a}-t}\right)$$

$h(x) = f(-ax)$ for $0 \leq x \leq 1$.

$$h''(x) = a^2 f''(-ax) = -\frac{a}{(1-x)(\bar{a}+xa)} \leq -a$$

and so

$$f(-ax) \leq -ax^2/2.$$

Doob Martingale

$\Omega = \{(A_1, A_2, \dots, A_n)\}$

Let $Z = Z(A_1, A_2, \dots, A_n)$ be a random variable with $E(Z) = 0$.

Define random variables, $X_0 = 0$ and

$$X_i = E(Z \mid A_1, A_2, \dots, A_i), \quad 1 \le i \le n$$

Claim

$X_0, X_1, \dots, X_n$ is a martingale w.r.t. $A_0, A_1, \dots, A_n$ with $X_0 = E(Z) = 0$ and $X_n = Z$.

Proof

$$E(X_{i+1} \mid A_0, A_1, \dots, A_i)$$

A_{i+1}

$$= E\big(E(Z \mid A_0, A_1, \dots, A_{i+1}) \mid A_0, A_1, \dots, A_i\big)$$

given $A_0, \dots, A_i$ we are averaging X_{i+1} over A_{i+1}

$$= X_i .$$

$\square$

Case 1

$$Z = Z_1 + Z_2 + \cdots + Z_n$$

where $Z_1, Z_2, \ldots, Z_n$ are independent.

Put $$X_i = \sum_{j=1}^{i} (Z_j - E(Z_j)).$$

$$X_{i+1} = X_i + Z_{i+1} - E(Z_{i+1})$$

$$E(X_{i+1} | X_i, X_{i-1}, \ldots X_1) = X_i$$

and all the derived inequalities apply.

In particular if $0 \leq Z_i \leq 1$ and $E(Z_i) = a_i$

then we get bounds on

$$\Pr\left(\left| Z - \sum a_i \right| \geq t \right)$$

by considering

$$\hat{Z}_i = Z_i - a_i \in [-a_i, 1 - a_i].$$

Case 2

$Z = Z(A_1, \ldots, A_n)$ and $A_0, A_1, \ldots, A_n$ are independent.

Theorem

If

$$a_i \leq Z(A_1, \ldots, A_i, \ldots A_n) - Z(A_1, \ldots, \hat{A_i}, \ldots, A_n) \leq b_i$$

for all $i, A_1, A_2, \ldots, A_n, \hat{A_i}$ then

$$\Pr(|Z - E(Z)| \geq t) \leq 2e^{-2t^2 / \sum_{i=1}^{n} (b_i - a_i)^2}.$$

We can assume w.l.o.g. that $E(Z) = 0$.

Now define $X_i = E(Z \mid A_1, \dots, A_i)$ as before

$X_{i+1} - X_i =$

$$\sum_{\hat{A}_{i+1}, \dots, \hat{A}_n} \underbrace{\boxed{\left(Z(A_1, \dots, A_{i+1}, \hat{A}_{i+2}, \dots, \hat{A}_n) - Z(A_1, \dots, A_i, \hat{A}_{i+1}, \dots, \hat{A}_n)\right)}}_{\in [a_i, b_i]} \cdot \Pr(\hat{A}_{i+1}) \cdot \Pr(\hat{A}_{i+2}) \cdot \ldots \cdot \Pr(\hat{A}_n)$$

So $a_i \leq X_{i+1} - X_i \leq b_i$.

$\square$

In $G_{n,p}$ we can take

(i) $A_1, A_2, \ldots, A_{\binom{n}{2}}$ as independent 0-1 random variables defining G.

(ii) $\tilde{A}_i = \{ (j, i) : j \leq i \text{ and } (j, i) \in E(G_{n,p}) \}$

Case 3

For $G_{n,m}$ we need a slight modification.

Suppose

$$Z = Z(u_1, u_2, \dots, u_m)$$

where $u_1, u_2, \dots, u_N$ is a random permutation of $\{1, 2, \dots, N\}$.

Suppose that

one interchange

$$a_i \le Z(u_1, \dots u_i, \dots, u_m) - Z(u_1, \dots, \hat{u}_i, \dots, u_m) \le b_i$$

then

$$\Pr(|Z - E(Z)| \ge t) \le 2e^{-t^2 / \sum_{i=1}^{m} (b_i - a_i)^2}$$

Now define $X_i = E[Z \mid u_1, u_2, \dots, u_i]$

$$X_{i+1} - X_i =$$

$$\sum_{\hat{u}_{i+1}, \dots, \hat{u}_N} \Big(Z(u_1, \dots, u_i, \hat{u}_{i+1}, \dots, \hat{u}_j = u_{i+1}, \dots \hat{u}_N) - Z(u_1, \dots u_i, u_{i+1}, \dots, \hat{u}_{i+1}, \dots \hat{u}_N) \Big) \times \frac{1}{(N-i)!}$$

B WE INTERCHANGE

$\square$