0.2 Inequalities for Sums of Bounded Independent

We start from Chernoff bound:

> [Theorem 2.1] Let $0<p<1$, let $X_1, \cdots, X_n$ be independent binary random variables, with $\Pr(X_k=1)=p$, $\Pr(X_k=0)=1-p$ for each k. Let $S_n=\sum_{k=1}^{n} X_k$, then for any $t \geq 0$,
> $$\Pr(|S_n-np| \geq nt) \leq 2e^{-2nt^2}$$

Pf: Recall that Markov's inequality: for $X \geq 0$ r.v.

$$\Pr(X \geq t) \leq E(X)/t \quad \text{for any } t>0.$$

Let $m=n(p+t)$, $h>0$. Then

$$\Pr(S_n \geq m) = \Pr(e^{hS_n} \geq e^{hm}) \overset{\text{by Markov's ineq.}}{\leq} e^{-hm}E(e^{hS_n}),$$

by the independence of X_k,

$$E(e^{hS_n}) = E\left(\prod_{k=1}^{n} e^{hX_k}\right) = \prod_{k=1}^{n} E(e^{hX_k}) = (1-p+pe^h)^n$$

Hence for any $h>0$

$$\Pr(S_n \geq m) \leq e^{-hm}(1-p+pe^h)^n$$

We may set $e^h = \frac{(p+t)(1-p)}{p(1-p-t)}$ to minimise the RHS.

then we get

$$\text{RHS} = \left(\frac{1-p}{1-p-t} \cdot \left(\frac{(p+t)(1-p)}{p(1-p-t)}\right)^{-(p+t)}\right)^n$$
$$= \left[\left(\frac{1-p}{1-p-t}\right)^{1-p-t}\left(\frac{p}{p+t}\right)^{p+t}\right]^n$$

Let q denote $1-p$, Let $f(t)=\ln\left(\left(\frac{q}{q-t}\right)^{q-t}\left(\frac{p}{p+t}\right)^{p+t}\right)$

$$\Rightarrow f(t) = (q-t)\ln\left(\frac{q}{q-t}\right) + (p+t)\ln\left(\frac{p}{p+t}\right)$$
$$f'(t) = \ln\frac{q-t}{q} + (q-t)\cdot\frac{q-t}{q}\cdot\frac{q}{(q-t)^2} + \ln\frac{p}{p+t} + (p+t)\cdot\frac{p+t}{p}\cdot\frac{-p}{(p+t)^2}$$
$$= \ln\left(\frac{p(q-t)}{q(p+t)}\right)$$

and $f''(t) = -\frac{1}{q-t} - \frac{1}{p+t} = -\frac{1}{(p+t)(1-(p+t))} \leq -4$

since $f(0)=f'(0)=0$, by Taylor's theorem that for $0 \leq t < q$,

$f(t) = \frac{t^2}{2}f''(s)$ for some s with $0 \leq s \leq t$

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

Hence $f(t) = -2t$, and we finish the proof that

$\Pr(S_n - np \geq nt) \leq e^{-2nt^2}$, the same to $\Pr(S_n - np \leq nt)$ □

An extension of the above theorem can be derived from following Lemma.

[Lemma 2.2] Let r.v. $X_1, \cdots, X_n$ be independent, with $0 \leq X_k \leq 1$ for each k. Let $S_n = \sum X_k$, let $\mu = E(S_n)$, let $p = \mu/n$ and $q = 1-p$. Then for any $0 \leq t \leq q$.

$$\Pr(S_n - \mu \geq nt) \leq \left(\left(\frac{p}{p+t}\right)^{p+t} \left(\frac{q}{q+t}\right)^{q-t} \right)^n$$

Pf. Let $m = \mu + nt$, let $p_k = E(X_k)$ for each k.

$$\Pr(S_n \geq m) = \Pr(e^{hS_n} \geq e^{hm}) \leq e^{-hm} E(e^{hS_n})$$

Since e^{hx} is convex, $e^{hx} \leq (1-x) + xe^h$ for $0 \leq x \leq 1$.

$$\begin{aligned} E(e^{hS_n}) &= E(e^{hS_{n-1}}) E(e^{hX_n}) \\ &\leq E(e^{hS_{n-1}})(1 - p_n + p_n e^h) \\ &\leq \prod_{k=1}^{n} (1 - p_k + p_k e^h) \\ &\leq \left(\sum_{k=1}^{n} (1 - p_k + p_k e^h) / n \right)^n \\ &= (1 - p + pe^h)^n \end{aligned}$$

$$\Rightarrow \Pr(S_n \geq m) \leq e^{-hm} (1 - p + pe^h)^n$$

Let $e^h = \frac{(p+t)(1-p)}{p(1-p-t)}$, we prove the Lemma □

Following theorem generalise Th 2.1 or improve when p is small).

[Theorem 2.3.] Let r.v. $X_1, \cdots, X_n$ be independent with $0 \leq X_k \leq 1$ for each k. Let $S_n = \sum_{k=1}^{n} X_k$, $\mu = E[S_n]$, let $p = \frac{\mu}{n}$, $q = 1-p$.

(a) For any $t \geq 0$, $Pr(|S_n - \mu| \geq nt) \leq 2e^{-2nt^2}$

(b) For any $\varepsilon > 0$.

$$Pr(S_n \geq (1+\varepsilon)\mu) \leq e^{-((1+\varepsilon)\ln(1+\varepsilon)-\varepsilon)\mu} \leq e^{-\frac{\varepsilon^2\mu}{2(1+\varepsilon/3)}}$$

(c) For any $\varepsilon > 0$,

$$Pr(S_n \leq (1-\varepsilon)\mu) \leq e^{-\frac{1}{2}\varepsilon^2\mu}.$$

Pf: (a) use Lemma 2.2 and follow the proof in Th.2.1.

(b) Due to Lemma 2.2. , let $t = \frac{\varepsilon}{n}\mu$, we have

$$Pr(S_n \geq (1+\varepsilon)\mu) \leq e^{-h(1+\varepsilon)\mu}(1-p+pe^h)^n$$

$$\leq e^{-h(1+\varepsilon)\mu} e^{\mu(e^h-1)} \quad \text{for any } h > 0$$

let $h = \ln(1+\varepsilon)$, we have

$$Pr(S_n \geq (1+\varepsilon)\mu) \leq e^{-[(1+\varepsilon)\ln(1+\varepsilon)-\varepsilon]\mu}$$

now we want to show $-(1+\varepsilon)\ln(1+\varepsilon) - \varepsilon \leq -\frac{3\varepsilon^2}{6+2\varepsilon}$

let $f_1(x) = (6+8x+2x^2)\ln(1+x) - 6x - 5x^2$

$$f_1'(x) = (8+4x)\ln(1+x) + (6+8x+2x^2)/(1+x) - 6 - 10x$$

$$= (8+4x)\ln(1+x) - 8x$$

$$f_1''(x) = 4\ln(1+x) + \frac{8+4x}{1+x} - 8$$

$$= 4\ln(1+x) + \frac{4}{1+x} - 4$$

$$f_1^{(3)}(x) = \frac{4}{1+x} - \frac{4}{(1+x)^2}$$

$$= \frac{4x}{(1+x)^2} \geq 0$$

$\Rightarrow f_1''(0) = 0$, $f_1''(x) \geq 0$

$\Rightarrow f_1'(0) = 0$, $f_1'(x) \geq 0$

$\Rightarrow f_1(0) = 0$, $f_1(0) \geq 0$

$$\Rightarrow Pr(S_n \geq (1+\varepsilon)\mu) \leq e^{-[(1+\varepsilon)\ln(1+\varepsilon)-\varepsilon]\mu} \leq e^{-\frac{\varepsilon^2\mu}{2(1+\varepsilon/3)}}$$

(c)

Let $Y_k = 1 - X_k$.

$$\Pr(S_n \le (1-\varepsilon)\mu) = \Pr(\ S_n - n + n - \mu \le -\varepsilon\mu)$$
$$= \Pr(\ \sum Y_k \ge n-\mu+\varepsilon\mu)$$

Due to Lemma 2.2. let $m = n-\mu+\varepsilon\mu =: n-\mu+n\cdot t'$, $t' = \frac{\varepsilon\mu}{n} = \varepsilon p$

let $p' = \frac{n-\mu}{n} = 1-p = q$, $q' = 1-p' = p$.

$$\Pr(\sum Y_k \ge m) \le \left(\left(\frac{p'}{p'+t'}\right)^{p'+t'}\left(\frac{q'}{q'-t'}\right)^{q'-t'}\right)^n$$
$$= \left(\left(\frac{q}{q+\varepsilon p}\right)^{q+\varepsilon p}\left(\frac{p}{p-\varepsilon p}\right)^{p-\varepsilon p}\right)^n$$

Let $f_1(x) = \ln\left(\left(\frac{q}{q+px}\right)^{q+px}\left(\frac{p}{p-px}\right)^{p-px}\right)$
$$= (q+px)\ln\left(\frac{q}{q+px}\right) - (p-px)\ln(1-x)$$

$$f_1'(x) = p\ln\left(\frac{q}{q+px}\right) + (q+px)\,\frac{q+px}{q}\cdot\frac{-pq}{(q+px)^2}$$
$$+ p\ln\left(\frac{1}{1-x}\right) + (p-px)\,\frac{1}{1-x}$$
$$= p\ln\left(\frac{q}{(q+px)(1-x)}\right)$$

$$f_1''(x) = -\frac{p^2}{q+px} + \frac{p}{1-x} = \frac{-p}{(1-x)(q+px)} \le -p$$

when $0<x<1$. then by the Taylor's theorem:

$$f_1(\varepsilon) = \frac{f_1''(s)}{2}\varepsilon^2 \quad \text{for some } s\in(0,\varepsilon).$$
$$\Rightarrow f_1(\varepsilon) \le -\frac{p}{2}\varepsilon^2.$$

$$\Rightarrow \Pr(\ S_n \le (1-\varepsilon)\mu) = \Pr(\ \sum Y_k \ge m)$$
$$\le e^{-\frac{np}{2}\varepsilon^2}$$
$$= e^{-\frac{1}{2}\varepsilon^2\mu} \qquad \square$$

We can genalise the bounds of X_k to $[a_k, b_k]$, i.e. $a_k \le X_k \le b_k$.

[Lemma 2.6] Let r.v. X satisfies $E(X)=0$ and $a \le X \le b$.
Then for any $h>0$, $E(e^{hX}) \le e^{\frac{1}{8}h^2(b-a)^2}$

Pf: Since e^{hx} gives a convex function of x. for $a\le x\le b$,

$$e^{hx} \le \frac{x-a}{b-a}e^{hb} + \frac{b-x}{b-a}e^{ha}$$
$$\Rightarrow E(e^{hX}) \le \frac{b}{b-a}e^{ha} - \frac{a}{b-a}e^{hb}$$

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

where $p = -\frac{a}{b-a}$, $1-p = \frac{b}{b-a}$, $y = (b-a)h$,

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

$$f'(y) = -p + \frac{pe^{y}}{1-p+pe^{y}} = -p + \frac{p}{p+(1-p)e^{-y}}$$

$$f''(y) = \frac{p(1-p)e^{-y}}{(p+(1-p)e^{-y})^2} \leq \frac{1}{4}$$

since $f(0)=0$, $f'(0)=0$, $f''(y) \leq \frac{1}{4}$ for $y>0$, by Taylor's th.

$$f(y) = \frac{1}{2}\cdot f''(s)y^2 \leq \frac{1}{8}y^2$$

$$\Rightarrow E(e^{hX}) \leq e^{\frac{1}{8}(b-a)^2h^2} \qquad \square$$

Hoeffding gives the following extension of Theorem 2.3 (a)

> **Theorem 2.5** Let r.v. $X_1, \cdots, X_n$ be independent, with $a_k \leq X_k \leq b_k$ for each k. Let $S_n = \sum X_n$ and $\mu = E(S_n)$, then for any $t \geq 0$,
> $$P(|S_n - \mu| \geq t) \leq 2e^{-2t^2/\sum(b_k-a_k)^2}$$

Pf: By Lemma 2.6, for $h>0$,

$$E(e^{h(S_n-\mu)}) = E\left(\prod_{k=1}^{n} e^{h(X_k - E(X_k))}\right)$$
$$= \prod_{k=1}^{n} E\left(e^{h(X_k - E(X_k))}\right)$$
$$\leq e^{\frac{1}{8}h^2\sum(b_k-a_k)^2}$$

Hence by Markov's inequality,

$$\Pr(S_n - \mu \geq t) \leq e^{-ht}E(e^{h(S_n-\mu)})$$
$$\leq e^{-ht+\frac{1}{8}h^2\sum(b_k-a_k)^2}$$

set $h = 4t/\sum(b_k-a_k)^2$, we obtain

$$\Pr(S_n - \mu \geq t) \leq e^{-2t^2/\sum(b_k-a_k)^2}$$

Finally, replace X by $-X$ to obtain

$$\Pr(S_n-\mu \le -t) \le e^{[illegible]}$$

which complete the proof $\square$

An other extension is under the condition where we know bounds of X_k and variance of X_k. We need following Lemma 2.8.

> [Lemma 2.8] Let $g(x) = \frac{1}{2} + \frac{1}{3!}x + \frac{x^2}{4!} + \cdots = (e^x - 1 - x)/x^2$, $x \neq 0$
>
> Then g is increasing; and if r.v. X s.t. $E(X)=0$ and $X \le b$, then $E(e^X) \le e^{g(b)\operatorname{Var}(X)}$

Pf: ① To show g is increasing, note that for $x \neq 0$,

$$g'(x) = x^{-3}\left((x-2)e^x + 2 + x\right)$$

and it is suffices to show $h(x) = (x-2)e^x + 2 + x \geq 0$

Now $h(0)=0$. $h'(x) = (x-1)e^x + 1$. $h'(0)=0$, $h''(x) = xe^x$,

so $h'(x) < 0$ for $x<0$, $h'(x) > 0$ for $x>0$, which implies

$h(x) \ge 0$ for all x. Thus g is increasing.

② note $e^x = 1 + x + x^2 g(x) \le 1 + x + x^2 g(b)$ for $x \le b$.

Hence, if $E(X)=0$ and $X \le b$, then

$$E[e^X] \le 1 + g(b)\operatorname{var}(X) \le e^{g(b)\operatorname{Var}(X)} \quad \square$$

The following results builds on work of Bernstein.

> [Theorem 2.7] Let r.v. $X_1, \cdots, X_n$ be independent, with $X_k - E[X_k] \le b$ for each k. Let $S_n = \sum X_k$, and let S_n have expected value μ and variance V. Then for any $t \geq 0$,
>
> $$\Pr(S_n - \mu \ge t) \le e^{-\left(\frac{V}{b^2}\right)\left((1+\varepsilon)\ln(1+\varepsilon) - \varepsilon\right)} \le e^{-\frac{t^2}{2V(1+(bt/3V))}}$$
>
> where $\varepsilon = bt/V$

Pf: $E(e^{h(S_n-\mu)}) = \prod E(e^{h(X_k - E(X_k))}) \overset{\text{Lemma 2.8}}{\leq} e^{g(hb)h^2 V}$

$$Pr(S_n-\mu \geq t) \leq e^{-ht} E(e^{h(S_n-\mu)}) \leq e^{-ht+g(hb)h^2 V} \quad (*)$$

To minimise this bound, let $f(h) = -ht + g(hb)h^2 V$

$$f'(h) = -t + g'(hb)\cdot bh^2 V + g(hb)\cdot 2hV$$

$$= -t + (hb)^{-3}((hb-2)e^{hb} + 2 + hb)\cdot bh^2 V + (hb)^{-2}(e^{hb} - 1 - hb)\cdot 2hV$$

$$= -t + b^{-1}e^{hb}V - b^{-1}V$$

let $f'(h) = 0$, we obtain $h = \frac{1}{b}\ln(1+\frac{bt}{V})$

Then $(*)$ implies

$$Pr(S_n-\mu \geq t) \leq e^{-\frac{t}{b}\ln(1+\frac{bt}{V}) + (\frac{bt}{V} - \ln(1+\frac{bt}{V}))\cdot\frac{V}{b^2}}$$

$$= e^{-\frac{V}{b^2}((1+\frac{bt}{V})\ln(1+\frac{bt}{V}) - \frac{bt}{V})}$$

By Lemma 2.4: For all $x \geq 0$,

$$(1+x)\ln(1+x) - x \geq 3x^2/(6+2x)$$

which is shown in Th 2.3 (b).

we have

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

Thus,

$$Pr(S_n-\mu \geq t) \leq e^{-\frac{V}{b^2}\cdot\frac{3(\frac{bt}{V})^2}{6+\frac{2bt}{V}}}$$

$$= e^{-\frac{t^2}{2(V+(bt/3V))}} \quad \square$$