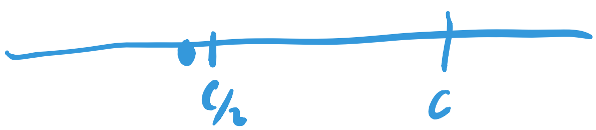
Chapter 2 Concentration Inequalities

If we want to upper bound

$$P\{X \geq c\} \leq P\{X - E[X] \geq \underbrace{c - E[X]}\} \leq P\{X - E[X] \geq c/2\}$$

- step 1 Bound $E[X]$
- step 2 Bound $P\{X - E[X] > t\}$

e.g. $E[X] \leq c/2$



Markov Inequality

If Y is a nonnegative r.v.

$$P\{Y \geq t\} \leq \frac{E[Y]}{t} \qquad (\text{all } t > 0)$$

Proof

$$P(Y \geq t) = E(1_{Y \geq t})$$
$$\leq E\left[\frac{Y}{t} 1_{Y \geq t}\right]$$
$$\leq E\left[\frac{Y}{t}\right]$$

Chebyshev inequality

$$P\{\,|S-E[S]|^2 \ge t^2\} \le \frac{Var(S)}{t^2}$$

Proof: Markov inequality for $Y=(S-E[S])^2$

e.g. $S = X_1 + \dots + X_n$ $\quad X_i \sim Be(\theta)$, independent

$Var(X_i) = \theta(1-\theta)$ $\qquad \hat{\theta} = \frac{S_n}{n}$

$Var(S) = n\theta(1-\theta)$ $\qquad E[\hat{\theta}] = \frac{E(S_n)}{n} = \theta$

$$P\left\{|\hat{\theta}-\theta| \ge t\sqrt{\frac{\theta(1-\theta)}{n}}\right\} \le \frac{\frac{\theta(1-\theta)}{n}}{t^2\,\frac{\theta(1-\theta)}{n}} = \frac{1}{t^2}$$

$$Var(\hat{\theta}) = \frac{\theta(1-\theta)}{n}$$

Central Limit Theorem

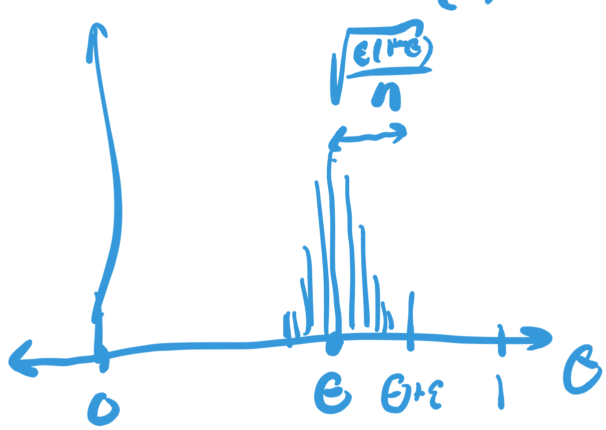
$$P\left(\frac{S_n - n\theta}{\sqrt{n\theta(1-\theta)}} \ge c\right) \longrightarrow Q(c) = \int_c^\infty \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\,du$$

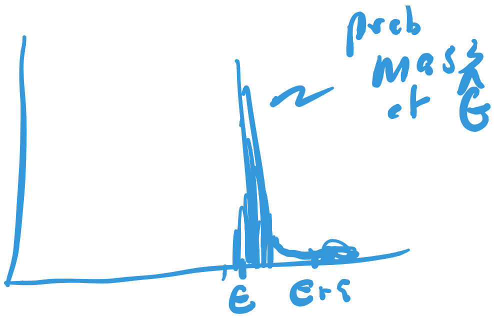
$$\le \int_c^\infty \frac{1}{\sqrt{2\pi}}\frac{u}{c} e^{-u^2/2}\,du$$

$$= \frac{1}{\sqrt{2\pi}\,c}\int_c^\infty u\,e^{-u^2/2}\,du$$

$$= \frac{e^{-c^2/2}}{\sqrt{2\pi}\,c}$$

$c > 0$





Chernoff Bound

$$P\{X-\underline{\underline{E(X)}}\geq t\}\leq E\left[e^{-s(t-(X-E(X)))}\right] \qquad s\geq 0$$

$$=\underbrace{e^{-st}E\left[\underline{e^{s(X-E(X))}}\right]}_{\text{min over } s \text{ to get tightest bound.}}$$

suppose X is $N(\mu,\sigma^2)$ wlog let $\mu=0$

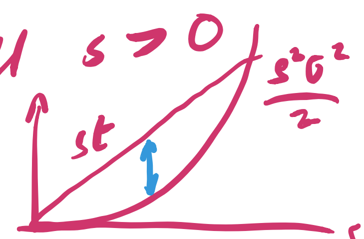
$$E[e^{sX}]=\int_{-\infty}^{\infty}e^{su}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-u^2/2\sigma^2}\,du$$

$$=\underbrace{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\left(\underbrace{\frac{u^2-2s\sigma^2u+s^2\sigma^4}{2\sigma^2}}_{\frac{(u-s\sigma^2)^2}{2\sigma^2}}\right)}du}_{1}\; e^{\frac{s^2\sigma^2}{2}}$$

so

$$P\{X\geq t\}\leq e^{-\left(st-\frac{s^2\sigma^2}{2}\right)}$$

all $s>0$



$\frac{d}{ds}()$ $\quad t-s\sigma^2=0 \quad s=\frac{t}{\sigma^2}$

$$\therefore \quad P\{X \geq t\} \leq e^{-t^2/2\sigma^2}$$

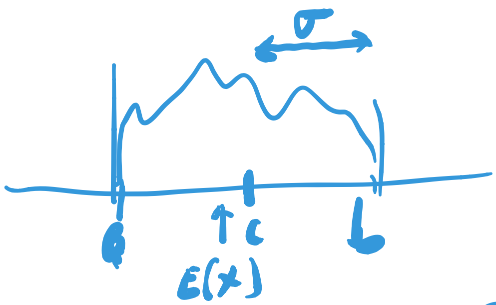
$$t = a\sigma \qquad P\{X \geq a\sigma\} \leq e^{-a^2/2}$$

Bounded random variables

Suppose X takes values in $[a,b]$

$$Var(X) = E(X - E[X])^2 \leq E[(X-c)^2]$$

for all c.



$$\leq \left(\frac{b-a}{2}\right)^2$$

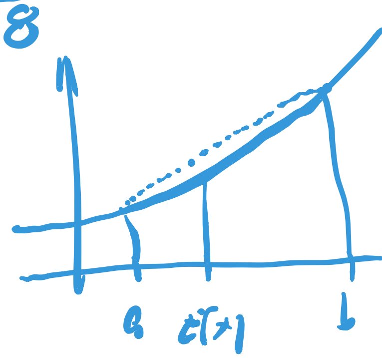
Take $c = \frac{a+b}{2}$

$$= \frac{(b-a)^2}{4}$$

$$\sigma = \sqrt{Var(X)} \leq \frac{b-a}{2}$$

Hoeffding lemma

$$E[e^{s(X - E[X])}] \leq e^{\frac{s^2(b-a)^2}{8}}$$



$$x \to e^{s(x - E(X))}$$

If X, Y independent

$$E[e^{s(X+Y)}] = E[e^{sX}]E[e^{sY}]$$

Hoeffding Thm

If $X_1, \dots X_n$ are independent

$X_i \in [a_i, b_i]$ (so X_i is subgaussian with scale parameter $v^2 = \frac{(b_i - a_i)^2}{4}$)

and $S_n = X_1 + \dots + X_n$ then

$$P\{ S_n - E(S_n) \geq t \} \leq \exp\left(\frac{2t^2}{\sum (b_i - a_i)^2} \right)$$

$$P\{ S_n - E(S_n) \leq -t \} \leq \quad "$$

$$P\{ |S_n - E(S_n)| \geq t \} \leq 2\exp\left(\quad " \quad \right)$$

McDiarmid inequality

Let $g : \mathbb{R}^n \to \mathbb{R}^n$ so that for $1 \leq i \leq n$

$$\left| g(x_1, \dots, x_i, \quad x_n) - g(x_1, \dots x_i' \dots x_n) \right| \leq c_i$$

"Bounded difference property with constants $c_1, \dots, c_n$."

and suppose $X_1, \dots, X_n$ are independent

Then $P\{g(X_1, \cdots X_n) - E[g(X_1, \cdots X_n)] \geq t\} \leq e^{-\frac{2t^2}{\sum_{i=1}^{n} c_i^2}}$

Proof Idea is to come up with another sequence of random variables $V_1, \cdots V_n$ such that $g(X_1, \cdots X_n) - E[g(X_1, \cdots X_n)] = V_1 + \cdots + V_n$

Consider sequence of predictions of $g(X_1, \cdots X_n)$ (assume $E[g(X_1, \cdots X_n)] = 0$)

$$
\begin{aligned}
&- E[g(X_1, \cdots X_n)] \quad \big\} V_1 \\
&- E[g(X_1, \cdots X_n) \mid X_1] \quad \big\} V_2 \\
&\quad \vdots \\
&\qquad\qquad\qquad\qquad\qquad\qquad \big] V_n \\
&- E[g(X_1, \cdots X_n) \mid X_1, X_2, \cdots X_n] = g(X_1, X_2 \cdots X_n)
\end{aligned}
$$

$$V_i = E[g(X_1, \cdots X_n) \mid X_1 \cdots X_i] - E[g(X_1, \cdots X_n) \mid X_1 \cdots X_{i-1}]$$

$\therefore$ bounded diff property

$$V_i \in [A_i(X_1, \cdots X_{i-1}), B_i(X_1, \cdots X_{i-1})]$$

with $B_i(X_1, \cdots X_{i-1}) - A_i(X_1, \cdots X_{i-1}) \leq c_i$

$$E\left[e^{s\,g(x_1 \cdots x_n)}\right]$$

$$= E\left[e^{s(V_1 + \cdots V_n)}\right]$$

$$= E\left[E\left(e^{s(V_1 \cdots + V_n)} \mid V_1, \cdots V_{n-1}\right]\right]$$

$$\leq E\left[e^{s(V_1 + \cdots + V_{n-1})}\right] e^{\frac{s^2 c_n^2}{8}}$$

$$\vdots$$

$$\leq e^{\frac{s^2}{8} \sum c_i^2}$$