EE6106 Lecture 6 (Date: 2 Feb 2024)

Recall: $\{X_i\} \sim$ iid, light tailed with mean μ

$$S_n = \sum_{i=1}^{n} X_i$$

$$\mathbb{P}(S_n > na) \leq e^{-n \sup_{s>0} [sa - \Lambda_X(s)]}$$

$a > \mu$

Chernoff bound

$$= \mathbb{P}\left(\frac{S_n}{n} - \mu > (a-\mu)\right)$$

can be used as the empirical estimator of μ

The learning algorithm not being able to know the underlying distribution is something to take note of.

HOEFFDING'S INEQUALITY

$\{X_i\} \sim$ iid with $a_i \leq X_i \leq b_i$ a.s.

independent (not necessarily i.i.d.)

$S_n = \sum_1^n X_i$, Then for $t > 0$,

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq t) \leq e^{\frac{-2t^2}{\sum (b_i - a_i)^2}}$$

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq t) \leq 2e^{\frac{-2t^2}{\sum (b_i - a_i)^2}}$$

Note that : Suppose we choose $a_i = a, b_i = b, t = ny$

$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geq ny) \leq e^{\frac{-2ny^2}{(b-a)^2}}$$

This is a bound over a FAMILY OF DISTRIBUTIONS, to be specific, those that are BOUNDED. We require no knowledge of the nature of the distribution.

Hoeffding's lemma

$X \sim$ support $[a, b]$, then for $s \in \mathbb{R}$,

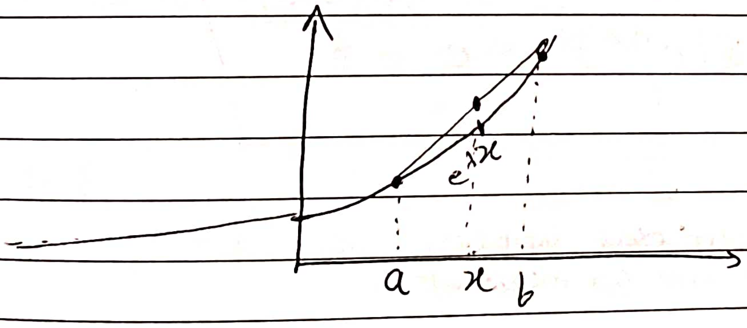
$$\mathbb{E}\left[e^{s(X - \mathbb{E}(X))}\right] \leq e^{\frac{s^2 (b-a)^2}{8}}$$

Remember the structure of a Gaussian MGF

Proof : WLOG, $\mathbb{E}[X] = 0$ (allowed since only the width of the support is of interest)

For $x \in [a, b]$,

$$e^{\lambda x} \leq \frac{b-x}{b-a} e^{\lambda a} + \frac{x-a}{b-a} e^{\lambda b} \quad \text{(convexity)}$$



$$\mathbb{E}[e^{\lambda x}] \leq \frac{b}{b-a} e^{\lambda a} - \frac{a}{b-a} e^{\lambda b}$$

$$= \left(\frac{b}{b-a}\right) e^{\lambda a} + \left(1 - \frac{b}{b-a}\right) e^{\lambda b}$$

$$= e^{\log\left[e^{\lambda a}\left(\frac{b}{b-a} + \left(1-\frac{b}{b-a}\right)e^{\lambda(b-a)}\right)\right]}$$

$$\leq \exp\left\{\lambda a + \left[\log\left(\frac{b}{b-a} + \left(1-\frac{b}{b-a}\right)\right)e^{\lambda(b-a)}\right]\right\}$$

$$= \exp\left\{\left(\frac{b}{b-a}-1\right)\lambda(b-a) + \log\left(\frac{b}{b-a} + \left(1-\frac{b}{b-a}\right)\right)e^{\lambda(b-a)}\right\}$$

let $\frac{b}{b-a} = z$, $\lambda(b-a) = y$

$$= \exp\{\underbrace{(z-1)y + \log(z + (1-z)e^y)}_{f(y)}\}$$

$f(0) = 0$

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

$f'(0) = 0$

HW $\quad f''(y) \leq \frac{1}{4} \; \forall y$

Then, using Taylor's Expansion, we have

$$f(y) \leq \frac{y^2}{8}$$

Now, we thus have

$$E[e^{\lambda x}] \leq \exp(f(y)) \leq e^{\frac{\lambda^2(b-a)^2}{8}}$$

$$\boxed{\therefore E[e^{\lambda x}] \leq e^{\frac{\lambda^2(b-a)^2}{8}}}$$

Universal upper bound on the MGF

Proof of Hoeffding's Inequality

$$P(S_n - E[S_n] \geq t) = P(e^{s[S_n - E[S_n]]} \geq e^{st}) \quad \text{where } s>0$$

M T W T F S S	
Page No.:	YOUVA
Date:	

$$\therefore \mathbb{P}(S_n - \mathbb{E}[S_n] \geq t) \leq e^{-st}\, \mathbb{E}\left[e^{s(S_n - \mathbb{E}(S_n))}\right]$$

$$= e^{-st} \prod_{i=1}^{n} \mathbb{E}\left[e^{s(X_i - \mathbb{E}(X_i))}\right]$$

$$\leq e^{-sb} \prod_{i=1}^{n} e^{\frac{s^2(b_i - a_i)^2}{8}}$$

$$\leq e^{-st}\, e^{\frac{s^2}{8}[\Sigma(b_i - a_i)^2]}$$

To minimize this, we can choose

$$s^* = \frac{4t}{\Sigma(b_i - a_i)^2}$$

$$\Rightarrow P(S_n - \mathbb{E}[S_n] \geq t) \leq e^{-\frac{2t^2}{\Sigma(b_i - a_i)^2}}$$

EXAMPLE : Given a coin with unknown bias p ($\mathbb{P}r\{Heads\} = p$)
Toss n times. H_n = # of heads
Then the natural estimate of the unknown bias p is

$$\hat{p} = \frac{H_n}{n}$$

Note this does not give us any sort of a confidence interval

$\mathbb{P}(|\hat{p} - p| \geq \varepsilon) \leq 2e^{-2n\varepsilon^2}$ ... using Hoeffding

To compute an error probability bound, we can replace the RHS by δ

Then the corresponding ϵ s.t. $\mathbb{P}(|\hat{p} - p| \geq \varepsilon) \leq \delta$ is

$$\epsilon = \sqrt{\frac{\log(2/\delta)}{2n}}$$

$\therefore$ with prob $1-\delta$,

$$|\hat{p} - p| \leq \sqrt{\frac{\log(2/\delta)}{2n}}$$

i.e. $$\underbrace{\hat{p} - \sqrt{\frac{\log(2/\delta)}{2n}}}_{\text{Lower Confidence (LCB)}} \leq p \leq \underbrace{\hat{p} + \sqrt{\frac{\log(2/\delta)}{2n}}}_{\text{Upper Confidence Bound (UCB)}}$$

Want these to be algorithm countable

One issue with Hoeffding-type inequalities is that they require the support of the random variables to be bounded.
Hence, we move on to:

$$\boxed{\text{" SUB-GAUSSIAN DISTRIBUTIONS"}}$$

$X \sim \mathcal{N}(\mu, \sigma^2) \rightarrow$ Normal or Gaussian

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Then the MGF of the Gaussian when $\mu = 0$ and $\sigma^2 = 1$ is

$$M_X(s) = e^{s^2/2}$$

Further if $\mu = 0$ and σ^2 is known,

$$M_X(s) = e^{\sigma^2 s^2/2}$$

HW Then, for $x \geq 0$, $P(X \geq x) \leq e^{-x^2/2\sigma^2}$ HW

$\downarrow$

$\mathcal{N}(0, \sigma^2)$

This intuitively means that the tail of the Gaussian is decaying even faster than the exponential distribution (?)

Def A zero mean random variable Z is σ-subGaussian if

$$M_Z(s) \leq e^{\sigma^2 s^2/2} \quad \forall s$$

Def $Z \sim \sigma$-subGaussian if $Z - E[Z] \sim \sigma$-subGaussian

Lemma $X \sim \sigma$-subGaussian with $E(X) = 0$
Then $P(X \geq x) \leq e^{-x^2/2\sigma^2}$

Proof: Use Chernoff bound approach (Markov + Exponentiation)

Page No.: YOUVA
Date:

HW Theorem
① $X \sim \sigma-\text{sub}G \Rightarrow Var(X) \leq \sigma^2$
② $X \sim \sigma-\text{sub}G \Rightarrow cX \sim |c|\sigma-\text{sub}G$
③ $X_1 \sim \sigma_1-\text{sub}G$, $X_2 \sim \sigma_2-\text{sub}G$, $X_1 \perp\!\!\!\perp X_2$, ⟶ independent
$\Rightarrow X_1 + X_2 \sim \sqrt{\sigma_1^2 + \sigma_2^2}-\text{sub}G$

Theorem Say $\{X_i\}$ are independent σ-sub G. Then for $t \geq 0$,
$$\mathbb{P}\left(\sum_{i=1}^{n}(X_i - E(X_i)) \geq t\right) \leq e^{-t^2/2n\sigma^2}$$
$$\mathbb{P}\left(\sum_{1}^{n}(X_i - E(X_i)) \leq -t\right) \leq e^{-\frac{t^2}{2n\sigma^2}}$$

Note: Structurally, this is almost identical to the Hoeffding inequality

Note: Gaussians are trivially sub-Gaussians

Note: In real-life, we know from the domain what the support of the distribution is going to be

Note: Verifying sub Gaussianity is not very easy in practice

Proof $\sum(X_i - E(X_i)) \sim \sigma\sqrt{n}$ SubG
$\therefore \mathbb{P}(\sum(X_i - E(X_i)) \geq t) \leq e^{-t^2/2n\sigma^2}$ □

IMP
1) Every bounded distribution is sub-Gaussian
2) Exponential distributions are NOT sub-Gaussian
(since sub-Gaussians have faster than doubly exponential tail decay)

HW : If X is 0 mean and has bounded support $[-B, B]$, then
$X \sim B-\text{sub}G$

HW : If X is 0 mean and has support $[a,b]$, then
$X \sim \left(\frac{b-a}{2}\right)\text{sub}G$

Next : Stochastic MAB