Lecture #2: concentration inequalities

Hoeffding lemma: For X a random variable s.t. $X \in [a,b]$ a.s., then $\forall s \in \mathbb{R}$:

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

Proof:

define for any $s \in \mathbb{R}$ $\quad \psi(s) = \ln \mathbb{E}\left[e^{sX}\right]$

Note that $\mathbb{E}\left[e^{sX}\right] = \int_\Omega e^{sX(\omega)}\, d\mathbb{P}(\omega)$ can be differentiated under the integral; for any compact I, $\forall s \in I, x \in [a,b]$, $\quad x e^{sx} \leq e^{\max_{s' \in I} |s'| \max(|a|,|b|)} \max(|a|,|b|)$

so: $$\psi'(s) = \frac{\mathbb{E}\left[X e^{sX}\right]}{\mathbb{E}\left[e^{sX}\right]}$$

Similarly: $$\psi''(s) = \frac{\mathbb{E}\left[X^2 e^{sX}\right]\mathbb{E}\left[e^{sX}\right] - \mathbb{E}\left[X e^{sX}\right]^2}{\mathbb{E}\left[e^{sX}\right]^2}$$

$$= \mathrm{Var}_{\mathbb{Q}}(X)$$

under the probability $\mathbb{Q}$ defined as:

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(\omega) = \frac{e^{sX(\omega)}}{\mathbb{E}_{\mathbb{P}}[e^{sX}]}$$

So $\psi''(s) = \inf_{\mu \in [0,1]} \mathbb{E}_{\mathbb{Q}}[(X-\mu)^2] \leq \mathbb{E}_{\mathbb{Q}}\left[\left(X - \frac{a+b}{2}\right)^2\right]$
$\leq \frac{(a-b)^2}{4}$,

ψ is $\mathcal{C}^2$, so that Taylor expansion yields for any $s \in \mathbb{R}$ and some $c_s \in [0, s]$:

$$\psi(s) = \psi(0) + s\psi(0) + \frac{s^2}{2}\psi''(c_s).$$

$$\leq 0 + s\,\mathbb{E}[X] + \frac{s^2}{8}.$$

$\square$

Hoeffding lemma (conditional version)

X r.v. such that $X \in [a, b]$ a.s. Then for all σ-algebra $\mathcal{G}$ and $s \in \mathbb{R}$: $\ln \mathbb{E}[e^{s(X - \mathbb{E}[X|\mathcal{G}])}|\mathcal{G}] \leq \frac{s^2}{8}(b-a)^2$

we could work with a similar (but adapted) proof → see exercise session #1

Let's prove it in the less elegant, but original way.

Proof Let $Y = X - \mathbb{E}[X|\mathcal{G}] \in [A, B]$

where
$A = a - \mathbb{E}[X|\mathcal{G}]$
$B = b - \mathbb{E}[X|\mathcal{G}]$
are both measurable
and $B - A = b - a \geq 0$

$$Y = \frac{B-Y}{B-A} A + \frac{Y-A}{B-A} B$$

since $y \mapsto e^{sy}$ is convex: $e^{sY} \leq \frac{B-Y}{B-A} e^{sA} + \frac{Y-A}{B-A} e^{sB}$

Taking $\mathbb{E}[\,|\mathcal{G}]$ using $\mathbb{E}[Y|\mathcal{G}] = 0$ and A, B $\mathcal{G}$-measurable

$$\mathbb{E}[e^{sY}|\mathcal{G}] \leq \frac{B}{B-A} e^{sA} - A e^{sB}$$

Now by a function study (eg the proof of unconditional Hoeffding lemma):

$\forall u, v \in \mathbb{R}, \forall p \in [0,1], \forall s \in \mathbb{R}$:

$$\ln\left(p e^{su} + (1-p) e^{sv}\right) \leq s\left(pu + (1-p)v\right) + \frac{s^2}{8}(v-u)^2$$

In particular:

$B \geq 0$
$A \leq 0$

$$\frac{B}{B-A} e^{sA} - \frac{A}{B-A} e^{sB} \leq \exp\left(s\left(\frac{BA}{B-A} - \frac{AB}{B-A}\right) + \frac{s^2}{8}(B-A)^2\right)$$

$$= \exp\left(\frac{\lambda^2}{8}(b-a)^2\right)$$

Finally,

$$\mathbb{E}[e^{\lambda Y} | \mathcal{G}] \leq \exp\left(\frac{\lambda^2}{8}(b-a)^2\right) \quad \square.$$

Hoeffding-Azuma inequality

Let $(\mathcal{F}_t)_{t \geq 0}$ be a filtration and let $(X_t)_{t \geq 1}$ be a sequence of adapted random variables (i.e $\forall t \geq 1$, X_t is $\mathcal{F}_t$-measurable), that are bounded $\forall t$, $a_t \leq X_t \leq b_t$ a.s. where $a_t, b_t \in \mathbb{R}$, then:

$$\forall \varepsilon > 0, \quad \mathbb{P}\left(\sum_{t=1}^{T} X_t - \mathbb{E}[X_t | \mathcal{F}_{t-1}] \geq \varepsilon\right) \leq \exp\left(-\frac{2\varepsilon^2}{\sum_{t=1}^{T}(b_t - a_t)^2}\right)$$

<ins>Note:</ins> Hoeffding's inequality is the special case when all X_t are independent and $\mathcal{F}_{t-1} = \sigma(X_1, \dots, X_{t-1})$, so that $\mathbb{E}[X_t | \mathcal{F}_{t-1}] = \mathbb{E}[X_t]$

<ins>**Proof**</ins>

Denote the martingale $S_T = \sum_{t=1}^{T} X_t - \mathbb{E}[X_t | \mathcal{F}_{t-1}]$

Markov inequality yields for any $\eta > 0$:

$$\mathbb{P}(S_T > \varepsilon) = \mathbb{P}(e^{\eta S_T} > e^{\eta\varepsilon}) \leq e^{-\eta\varepsilon}\, \mathbb{E}[e^{\eta S_T}]$$

We show by induction that $\mathbb{E}[e^{\eta S_T}] \leqslant \exp\left(\frac{\eta^2}{8}\sum_{t=1}^{T}(b_t - a_t)^2\right)$

- For $T=1$, conditional Hoeffding's lemma gives:

$$\mathbb{E}[e^{\eta S_1} | \mathcal{F}_0] \leqslant \exp\left(\frac{\eta^2}{8}\sum_{t=1}^{T}(b_t - a_t)^2\right)$$

so same goes for $\mathbb{E}[e^{\eta S_1}]$.

- Assume it holds for $T \geqslant 1$.

$$\mathbb{E}[e^{\eta S_{T+1}}] = \mathbb{E}\left[\mathbb{E}[e^{\eta S_{T+1}} | \mathcal{F}_T]\right]$$

$$= \mathbb{E}\left[\mathbb{E}\left[e^{\eta S_T} e^{\eta(X_{T+1} - \mathbb{E}[X_{T+1}|\mathcal{F}_T])} \middle| \mathcal{F}_T\right]\right]$$

$$= \mathbb{E}\left[e^{\eta S_T}\ \mathbb{E}\left[e^{\eta(X_{T+1} - \mathbb{E}[X_{T+1}|\mathcal{F}_T])} \middle| \mathcal{F}_T\right]\right]$$

$$\leqslant \mathbb{E}[e^{\eta S_T}] \quad e^{\frac{\eta^2}{8}(b_{T+1} - a_{T+1})^2}$$

↑ conditional Hoeffding lemma

$$\leqslant e^{\frac{\eta^2}{8}\sum_{t=1}^{T}(b_t - a_t)^2} \quad e^{\frac{\eta^2}{8}(b_{T+1} - a_{T+1})^2}$$

↑ induction hypothesis

So previous Markov inequality becomes

$$\mathbb{P}(S_T \geqslant \varepsilon) \leqslant e^{-\eta\varepsilon + \frac{\eta^2}{8}\sum_{t=1}^{T}(b_t - a_t)^2}$$

This holds for any $\eta > 0$. Taking $\eta = \frac{4\varepsilon}{\sum_{t=1}^{T}(b_t - a_t)^2}$ minimises the right term so that:

$$P(S_T \geq \varepsilon) \leq e^{-\frac{2\varepsilon^2}{\sum_{t=1}^{T}(b_t - a_t)^2}} \qquad \square.$$

Hoeffding (-Azuma) inequality is very useful in stochastic bandits problems (next lectures).

Yet all the above inequalities hold for bounded random variables.

Can we have similar versions with unbounded variables?

Definition (sub-Gaussian variables)

A r.v. $X \in \mathbb{R}$ is σ sub-Gaussian if $\forall s \in \mathbb{R}$, $\mathbb{E}\left[e^{s(X - \mathbb{E}[X])}\right] \leq \exp\left(\frac{\sigma^2 s^2}{2}\right)$.

i.e. if it satisfies Hoeffding's lemma!

Examples:

- if $X \in [a, b]$ a.s., then it is $\frac{(b-a)^2}{4}$ sub-Gaussian
- if $X \sim \mathcal{N}(\mu, \sigma^2)$, it is σ sub-Gaussian.

Indeed, assume $\mu = 0$, $\sigma = 1$

we have

$$\mathbb{E}[e^{sX}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2} + sx} \, dx$$

$$= \frac{e^{\frac{s^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x-s)^2} \, dx = e^{\frac{s^2}{2}} \quad \square$$

The general case μ, σ is the same by rescaling argument.

Hoeffding's inequality holds for σ sub-Gaussian random variables:

Hoeffding inequality (sub-Gaussian version)

Let $(X_t)_{t \geq 1}$ be independent random variables, where each X_t is σ_t sub-Gaussian. Then for any $\varepsilon > 0$:

$$\mathbb{P}\left(\sum_{t=1}^{T} X_t - \mathbb{E}[X_t] \geq \varepsilon \right) \leq e^{-\frac{\varepsilon^2}{2\sum_{t=1}^{T} \sigma_t^2}}$$

- Same proof as in the bounded case.
- Hoeffding-Azuma can be extended to the sub-Gaussian case, when carefully handling the assumptions.

see exercise session #1 for a sub-Gaussian Hoeffding-Azuma inequality

Bernstein Inequality

Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration and let $(X_t)_{t\geq 1}$ be a sequence of adapted random variables (i.e $\forall t\geq 1$, X_t is $\mathcal{F}_t$-measurable), that are bounded $\forall t$, $0\leq X_t\leq 1$ a.s. and $\mathrm{Var}(X_t|\mathcal{F}_{t-1})\leq \sigma_t^2$, where $a_t, b_t, \sigma_t \in \mathbb{R}$. Then

$$\forall \varepsilon>0, \quad \mathbb{P}\Big(\sum_{t=1}^T X_t - \mathbb{E}[X_t|\mathcal{F}_{t-1}] \geq \varepsilon\Big) \leq \exp\Big(\frac{-\varepsilon^2}{2\sum_{t=1}^T \sigma_t^2 + \frac{2}{3}\varepsilon}\Big).$$

Comparison with Hoeffding-Azuma: $\exp\left(-\frac{2\varepsilon^2}{T}\right)$ (with $a_t=0, b_t=1$)

With only $0\leq X_t\leq 1$, we can only guarantee $\mathrm{Var}(X_t|\mathcal{F}_{t-1})\leq \frac{1}{4}$, in which case Hoeffding-Azuma is better (no additional ε term below)

But if $\sigma_t^2 \ll \frac{1}{4}$, Bernstein can be much better.

→ this is for example useful when $X_t \sim \mathrm{Ber}(p_t)$ with a small (or large) p_t.

- often used with ε scaling in T, while Hoeffding is used with ε scaling in $\sqrt{T}$

Proof: Similar to Hoeffding-Azuma proof, but using

Bernstein Lemma: Let X be a random variable in $[0,1]$. Then for any $\eta>0$:

$$\ln(\mathbb{E}[e^{\eta X}]) \leq \eta\mathbb{E}[X] + (e^\eta - \eta - 1)\mathrm{Var}(X)$$

see exercise at end of lecture

Denote the martingale $S_T = \sum_{t=1}^{T} X_t - \mathbb{E}[X_t | \mathcal{F}_{t-1}]$

let $\eta > 0$

$$\mathbb{P}(S_T \geq \varepsilon) \leq e^{-\eta\varepsilon} \mathbb{E}[e^{\eta S_T}]$$

We can again show by induction $\mathbb{E}[e^{\eta S_T}] \leq e^{(e^\eta - \eta - 1)\sum_{t=1}^{T} \sigma_t^2}$

So that

$$\mathbb{P}(S_T \geq \varepsilon) \leq e^{(e^\eta - \eta - 1)\underbrace{\sum_{t=1}^{T} \sigma_t^2}_{V} - \eta\varepsilon} \quad \text{for any } \eta > 0.$$

Minimizing the quantity $\underbrace{(e^\eta - \eta - 1) V - \eta \varepsilon}_{f(\eta)}$:

$f'(\eta) = 0.$

$$V e^\eta = V + \varepsilon.$$

$$\eta^* = \ln\left(\frac{V+\varepsilon}{V}\right)$$

and $f(\eta^*) = V + \varepsilon - (V+\varepsilon)\ln\left(1 + \frac{\varepsilon}{V}\right) - V.$

so: $\mathbb{P}(S_T \geq \varepsilon) \leq \exp\left(-h\left(\frac{\varepsilon}{\sum_{t=1}^{T}\sigma_t^2}\right) \sum_{t=1}^{T}\sigma_t^2\right)$ Bennett's inequality.

where $h(u) = (1+u)\ln(1+u) - u$.

see below

Using $h(u) \geq \dfrac{u^2}{2+\frac{2u}{3}}$ for $u \geq 0$ finally yields

$$\mathbb{P}(S_T \geq \varepsilon) \leq \exp\left(-\frac{\varepsilon^2}{2\sum_{t=1}^{T}\sigma_t^2 + \frac{2\varepsilon}{3}}\right) \square.$$

Lemma For $h(u) = (1+u)\ln(1+u) - u$

$$\forall u \geq 0, \quad h(u) \geq \frac{u^2}{2+\frac{2u}{3}}.$$

Proof. (omitted in class) This is simply by comparison of the derivatives.

Define $f(u) = h(u) - \dfrac{u^2}{2+\frac{2}{3}u}$

$$f'(u) = \ln(1+u) - \frac{2u\left(2+\frac{2}{3}u\right) - \frac{2}{3}u^2}{\left(2+\frac{2}{3}u\right)^2}$$

$$= \ln(1+u) - \frac{4u + \frac{2}{3}u^2}{\left(2+\frac{2}{3}u\right)^2}.$$

$$f''(u) = \frac{1}{1+u} - \frac{\left(4+\frac{4}{3}u\right)\left(2+\frac{2}{3}u\right) - \left(4+\frac{4}{3}u\right)\left(4u+\frac{2}{3}u^2\right)}{\left(2+\frac{2}{3}u\right)^3}$$

$$= \frac{1}{1+u} - 2\,\frac{2+\frac{2}{3}u - 4u - \frac{2}{3}u^2}{\left(2+\frac{2}{3}u\right)^2}$$

$$= \frac{1}{1+u} - 2\,\frac{2 - \frac{10}{3}u - \frac{2}{3}u^2}{\left(2+\frac{2}{3}u\right)^2}$$

$$= \frac{\left(2+\frac{2}{3}u\right)^2 - (1+u)\left(4 - \frac{20}{3}u - \frac{4}{3}u^2\right)}{(1+u)\left(2+\frac{2}{3}u\right)^2} = \frac{0 + \left(\frac{8}{3}+\frac{20}{3}-4\right)u + \left(\frac{4}{9}+\frac{20}{3}+\frac{4}{3}\right)u^2 + \frac{4}{3}u^3}{(1+u)\left(2+\frac{2}{3}u\right)^2}$$

≥ 0 for $u \geq 0$

$f(0) = 0$

$f'(0) = 0$ $\quad$ f convex on $\mathbb{R}_+$

$\rightsquigarrow$ $f' \geq 0$ on $\mathbb{R}_+$

$f \geq 0$ on $\mathbb{R}_+$.

Let X be a centered random variable in $\mathbb{R}$. Show that $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4 \Rightarrow 5$.

1. *Laplace transform:* for any $\eta \in \mathbb{R}$, $\ln(\mathbb{E}[e^{\eta X}]) \leq \frac{\sigma^2\eta^2}{2}$;
2. *Concentration:* for any $\varepsilon > 0$, $\max\{\mathbb{P}(X \geq \varepsilon), \mathbb{P}(X \leq -\varepsilon)\} \leq \exp(\frac{-\varepsilon^2}{2\sigma^2})$;
3. *Moment condition:* for any $q \in \mathbb{N}^*$, $\mathbb{E}[X^{2q}] \leq 2q!(2\sigma^2)^q$;
4. *Orlicz condition:* $\mathbb{E}[\exp(\frac{X^2}{4\sigma^2})] \leq 4$;
5. *Laplace transform:* for any $\eta \in \mathbb{R}$, $\ln(\mathbb{E}[e^{\eta X}]) \leq \frac{20\sigma^2\eta^2}{2}$.

1 sub-Gaussian variable

Prove Bernstein Lemma

Bernstein Lemma: Let X be a random variable in $[0, 1]$. Then for any $\eta > 0$:

$$\ln(\mathbb{E}[e^{\eta X}]) \leq \eta \mathbb{E}[X] + (e^{\eta} - \eta - 1)\,\mathrm{Var}(X)$$

Hint: Consider the function $\varphi : x \mapsto \frac{e^{-x}-x-1}{x^2}$, which is non-decreasing in $\mathbb{R}_+$

Solutions

Solution: 1) $\implies$ 2) is Hoeffding inequality for a single random variable.

For 2) $\implies$ 3),

$$\begin{aligned}
\mathbb{E}[X^{2q}] &= \int_0^{+\infty} \mathbb{P}(X^{2q} > u)\mathrm{d}u \\
&= \int_0^{+\infty} \mathbb{P}(|X| > u^{\frac{1}{2q}})\mathrm{d}u \\
&\leq 2\int_0^{+\infty} \exp(\frac{-u^{1/q}}{2\sigma^2})\mathrm{d}u \\
&= (2\sigma^2)^q 2q \int_0^{+\infty} \exp(-v)v^{q-1}\mathrm{d}v & v = \frac{u^{1/q}}{2\sigma^2} \\
&= (2\sigma^2)^q 2q\Gamma(q) \\
&= 2(2\sigma^2)^q q!
\end{aligned}$$

For 3) $\implies$ 4), the monotone convergence theorem gives

$$\mathbb{E}[\exp(\frac{X^2}{4\sigma^2})] = \mathbb{E}[\sum_{k=0}^{\infty} \frac{X^{2k}}{(2\sigma^2)^k k!}\frac{1}{2^k}] \leq 2\sum_{k=0}^{\infty}\frac{1}{2^k} = 4.$$

For 4) $\implies$ 5), using the fact that X is centered, we have for any $\eta \in \mathbb{R}$

$$\begin{aligned}
\mathbb{E}[\exp(\eta X)] &= \mathbb{E}[\sum_{k=0}^{\infty}\frac{(\eta X)^k}{k!}] \\
&= 1 + \mathbb{E}[\sum_{k=2}^{\infty}\frac{(\eta X)^k}{k!}] \\
&\leq 1 + \frac{\eta^2}{2}\mathbb{E}[X^2\exp(|\eta X|)] \\
&\leq 1 + \frac{\eta^2}{2}\exp(2\sigma^2\eta^2)\mathbb{E}[X^2\exp(\frac{X^2}{8\sigma^2})] & \inf_a(\frac{\eta^2}{2a} + \frac{aX^2}{2}) = \eta|X|, a = \frac{1}{4\sigma^2} \\
&\leq 1 + 2\sigma^2\eta^2\exp(2\sigma^2\eta^2)\mathbb{E}[\exp(\frac{X^2}{4\sigma^2})] & z \leq \exp(\frac{z}{2}) \\
&\leq (1 + 8\sigma^2\eta^2)\exp(2\sigma^2\eta^2) \leq \exp\left(\frac{20\sigma^2\eta^2}{2}\right) & 1 + z \leq e^z.
\end{aligned}$$

Bernstein

φ ← defined in 0 by continuity.

φ is non-decreasing, so: $\varphi(\eta(X-\mathbb{E}[X])) \leqslant \varphi(\eta)$ a.s.

ie $$e^{\eta(X-\mathbb{E}[X])} - \eta(X-\mathbb{E}[X]) - 1 \leqslant \cancel{\eta^2}(X-\mathbb{E}[X])^2 \quad \frac{e^\eta-\eta-1}{\cancel{\eta^2}}$$

taking $\mathbb{E}$: $$\mathbb{E}[e^{\eta X}]\, e^{-\eta \mathbb{E}[X]} \leqslant 1 + \mathrm{Var}(X)\,(e^\eta-\eta-1)$$

Taking $\ln$ $$\ln\left(\mathbb{E}[e^{\eta X}]\right) \leqslant \ln\left(1+\mathrm{Var}(X)(e^\eta-\eta-1)\right) + \eta\,\mathbb{E}[X].$$

$\ln(u+1)\leqslant u$ $$\leqslant \mathrm{Var}(X)(e^\eta-\eta-1) + \eta\,\mathbb{E}[X].$$

In this exercise, we are trying to prove the conditional Hoeffding's lemma with a similar proof technique we used for the Hoeffding lemma (without conditioning). Consider a random variable X such that $X \in [a,b]$ almost surely and a σ-algebra $\mathcal{G}$. Define the function $\psi : s \mapsto \ln(\mathbb{E}[e^{sX}] \mid \mathcal{G})$.

1) Justify that ψ is twice continuously differentiable on $\mathbb{R}$ and that for any $s \in \mathbb{R}$:

$$\psi'(s) = \frac{\mathbb{E}[Xe^{sX} \mid \mathcal{G}]}{\mathbb{E}[e^{sX} \mid \mathcal{G}]}$$

$$\psi''(s) = \frac{\mathbb{E}[X^2 e^{sX} \mid \mathcal{G}]\mathbb{E}[e^{sX} \mid \mathcal{G}] - \left(\mathbb{E}[Xe^{sX} \mid \mathcal{G}]\right)^2}{\left(\mathbb{E}[e^{sX} \mid \mathcal{G}]\right)^2}.$$

2) Show that we can define the probability distribution $\mathbb{Q}_s$ as

$$\frac{d\mathbb{Q}_s}{d\mathbb{P}} = \frac{e^{sX}}{\mathbb{E}[e^{sX} \mid \mathcal{G}]}$$

where $X \sim \mathbb{P}$.

3) Show that for any random variable Z, we have

$$\mathbb{E}_{\mathbb{P}}[Z\frac{e^{sX}}{\mathbb{E}[e^{sX} \mid \mathcal{G}]} \mid \mathcal{G}] = \mathbb{E}_{\mathbb{Q}_s}[Z \mid \mathcal{G}].$$

4) Deduce that

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