

## Course 1 : Probability toolbox

### o). Central limit theorem

$$f: \mathbb{R} \rightarrow \mathbb{R}$$

$$x_1, \dots, x_m \text{ iid s.t } \text{Var}[f(x_i)] < +\infty$$

$$\frac{1}{m} \sum_{i=1}^m f(x_i) = \mathbb{E} f(x_1) + \left[ \frac{\text{Var}[f(x_1)]}{m} \right]^{\frac{1}{2}} Z_m$$

$$\text{with } Z_m \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Assume  $f$  L-Lipschitz

$x_i$  finite variance  $\sigma^2$

$$\text{Then } \text{Var}[f(x_1)] = \frac{1}{2} \mathbb{E}[(f(x_1) - f(x_2))^2]$$

$$\leq \frac{L^2}{2} \mathbb{E}[(x_1 - x_2)^2]$$

$$\text{Var}[f(x_1)] \leq L^2 \sigma^2$$

We deduce that for  $n > 0$

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left[ \frac{1}{n} \sum_{i=1}^n f(x_i) - \mathbb{E} f(x_i) \geq \frac{\sigma}{\sqrt{n}} \right]$$

$$\leq \mathbb{P}(Z \geq z) \quad \text{with } Z \sim \mathcal{N}(0, 1)$$

$$\leq e^{-\frac{z^2}{2}}$$

Indeed  $h(n) := e^{-\frac{n^2}{2}} - \frac{1}{\sqrt{2\pi}} \int_n^\infty e^{-\frac{t^2}{2}} dt$

$$1) h'(n) = e^{-\frac{n^2}{2}} \left[ \frac{1}{\sqrt{2\pi}} - n \right] \quad \text{which is} \geq 0 \text{ for } n \leq \frac{1}{\sqrt{2\pi}}$$

$$2) h(0) = 1 - \frac{1}{2} > 0$$

$$1+2 : h > 0 \text{ for } n \in [0, \frac{1}{\sqrt{2\pi}}]$$

$$3) \text{ for } n > \frac{1}{\sqrt{2\pi}} : \frac{1}{\sqrt{2\pi}} \int_n^\infty e^{-\frac{t^2}{2}} dt \leq \int_n^{+\infty} t e^{-\frac{t^2}{2}} dt = e^{-\frac{n^2}{2}} \square$$

Goal Give a non asymptotic version of this result.

## I] From Markov to Chernoff

Markov  $X \geq 0$ ,  $\Pr[X \geq t] \leq \frac{\mathbb{E} X}{t}$ , for all  $t > 0$

Chebyshev Apply Markov to  $(X - \mu)^2$  with  $\mu = \mathbb{E} X$

$\Pr[|X - \mu| \geq t] \leq \frac{\text{Var } X}{t^2}$ , for all  $t > 0$

Order  $k$  Apply Markov to  $|X - \mu|^k$

①  $\Pr[|X - \mu| \geq t] \leq \frac{\mathbb{E} |X - \mu|^k}{t^k}$ , for all  $t > 0$

Chernoff bound

Moment Generating Function

$\exists b > 0$ ,  $\forall \lambda \in [-b, b]$ ,  $\Phi(\lambda) = \mathbb{E}[e^{\lambda(X-\mu)}] < \infty$

②  $\Pr[X - \mu \geq t] = \Pr[e^{\lambda(X-\mu)} \geq e^{\lambda t}] \leq \frac{\Phi(\lambda)}{e^{\lambda t}}$ ,  $\forall \lambda \in [0, b]$

Optimizing  $\lambda$ :

$$\log \Pr[X - \mu \geq t] \leq \inf_{\lambda \in [0, b]} \left\{ \log[\Phi(\lambda)] - \lambda t \right\}$$

Remark: Suppose  $X \geq 0$ , rt.

$$C := \inf_{k \geq 0} \frac{\mathbb{E}|X|^k}{s^k}$$

$$\begin{aligned} \text{then } e^{\lambda s} &= \sum_{k \geq 0} \frac{\lambda^k}{k!} s^k \\ &\leq C \sum_{k \geq 0} \frac{\lambda^k}{k!} \mathbb{E}[|X|^k] \\ &= C \mathbb{E} e^{\lambda X} \end{aligned}$$

$$\text{hence } \inf_{k \geq 0} \frac{\mathbb{E}[|X|^k]}{s^k} \leq \inf_{\lambda \geq 0} \frac{\mathbb{E} e^{\lambda X}}{e^{\lambda s}}$$

→ the moment bound ① with an optimal choice of  $k$  is never worse than ②

## II) Sub-Gaussian and Hoeffding

The gaussian case:  $X \sim N(\mu, \sigma^2)$

$$\Psi_X(\lambda) = \mathbb{E}[e^{\lambda(X-\mu)}] = e^{\frac{\sigma^2\lambda^2}{2}}, \text{ for all } \lambda \in \mathbb{R}$$

$$\inf_{\lambda \geq 0} \left\{ \log \Psi_X(\lambda) - \lambda \epsilon \right\} = \inf_{\lambda \geq 0} \left\{ \frac{\lambda^2 \sigma^2}{2} - \lambda \epsilon \right\} = -\frac{\epsilon^2}{2\sigma^2}$$

We uncover:

$$\boxed{\mathbb{P}[X - \mu \geq \epsilon] \leq e^{-\frac{\epsilon^2}{2\sigma^2}}}$$

Def (Sub-Gaussian)

A r.v  $X$  with mean  $\mu = \mathbb{E}[X]$  is Sub-Gaussian if there exists  $\sigma > 0$  s.t

$$\mathbb{E}[e^{\lambda(X-\mu)}] \leq e^{\frac{\sigma^2\lambda^2}{2}}, \text{ for all } \lambda \in \mathbb{R}$$

$\sigma$  is referred to as the sub-gaussian parameter

example  $\varepsilon \sim \text{Unif}(\{-1, +1\})$   
prove that  $\sigma = 1$

example Bounded R.V. :  $X \in [a, b]$

$$\Psi_x(1) = \mathbb{E}_x [e^{\lambda(x - \mathbb{E}_x[x])}] \quad | X \text{ ind. copy}$$

$$\leq \mathbb{E}_{x, x'} [e^{\lambda(x - x')}]$$

$$= \mathbb{E}_{x, x', \varepsilon} [e^{\lambda \varepsilon (x - x')}] \quad | \begin{array}{l} \varepsilon \text{ ind} \\ \varepsilon \sim \text{Unif}(\{-1, +1\}) \end{array}$$
$$[(x - x') \sim \varepsilon(x - x')]$$

$$\leq \mathbb{E}_{x, x'} [e^{\lambda^2 \frac{(x - x')^2}{2}}] \quad \text{using previous example}$$

$$|x - x'| \leq b - a \quad \leq e^{\lambda^2 \frac{(b - a)^2}{2}}$$

$$\rightarrow \boxed{\sigma = b - a}$$

one can prove:

$$\boxed{\sigma = \frac{b - a}{2}}$$

$\rightarrow$  We have used a "symmetrization" argument.

Prop  $X_1, X_2$  ind. sub-Gaussian with parameters  $\sigma_1, \sigma_2$  then  $X_1 + X_2$  is sub-Gaussian with parameter  $\sqrt{\sigma_1^2 + \sigma_2^2}$ .

proof:

$$\begin{aligned} \varphi_{X_1+X_2}(\lambda) &= \mathbb{E} e^{\lambda(X_1+X_2-\mu_1-\mu_2)} \\ &\stackrel{\text{ind}}{=} \mathbb{E} e^{\lambda(X_1-\mu_1)} \mathbb{E} e^{\lambda(X_2-\mu_2)} \\ &\leq e^{\frac{\lambda^2}{2}(\sigma_1^2 + \sigma_2^2)} \quad \square \end{aligned}$$

Prop (Hoeffding bound)

If  $X_i, i=1, \dots, n$  are independent  $X_i$  has mean  $\mu_i$  and sub-Gaussian parameter  $\sigma_i$ .

Then  $\forall \epsilon \geq 0$ ,

$$P\left[\sum_{i=1}^n (X_i - \mu_i) \geq \epsilon\right] \leq \exp\left[-\frac{\epsilon^2}{2 \sum_{i=1}^n \sigma_i^2}\right]$$

Cor:  $X_i \in [a, b]$  independent

$$P\left[\sum_{i=1}^n (X_i - \mu_i) \geq \epsilon\right] \leq e^{-\frac{\epsilon^2}{m(b-a)^2}}, \quad \forall \epsilon \geq 0$$

### III] Sub-exponential

Def (Sub-exponential)

A r.v.  $X$  with mean  $\mu$  is sub-exponential

if there are non-negative parameters  $(\beta, \alpha)$  s.t.

$$\mathbb{E}[e^{\lambda(X-\mu)}] \leq e^{\beta \frac{\lambda^2}{2}}, \text{ for all } |\lambda| < \frac{1}{\alpha}$$

example  $Z \sim \mathcal{N}(0,1)$  and  $X = Z^2$

One can check that:

$$\begin{aligned} \cdot \mu &= \mathbb{E}X = 1 \\ \cdot \varphi_x(\lambda) &= \mathbb{E}[e^{\lambda(X-1)}] \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\lambda(z^2-1)} e^{-\frac{z^2}{2}} dz \\ &= \frac{e^{-\lambda}}{\sqrt{1-2\lambda}} \end{aligned}$$

so that  $\varphi_x(\lambda) = \infty$  for  $\lambda \geq \frac{1}{2}$

$\rightarrow X$  is not sub-Gaussian

$$\text{But } \frac{e^{-\lambda}}{\sqrt{1-2\lambda}} < e^{-2\lambda^2} = e^{4\lambda - \frac{\lambda^2}{2}} \quad \forall |\lambda| \leq \frac{1}{4}$$

$\rightarrow X$  is sub-exponential  $(V, \alpha) = (2, 4)$

Prop (Sub-exponential tail bound)

$X$  sub-exponential  $(V, \alpha)$

$$P[X - \mu \geq t] \leq \begin{cases} e^{-\frac{t^2}{2\alpha^2}} & \text{if } 0 \leq t \leq \frac{V^2}{2} \\ e^{-\frac{t}{2\alpha}} & \text{if } t > \frac{V^2}{2} \end{cases}$$

Proof Wlog,  $\mu = 0$

$$\text{Chernoff: } P[X \geq t] \leq \underbrace{\frac{e^{\lambda^2 \alpha^2}}{e^{\lambda t}}}_{g(\lambda, t)}, \quad \forall \lambda \in [0, \frac{1}{2}]$$

Fix  $t \geq 0$ , compute  $g^*(t) = \inf_{\lambda \in [0, \frac{1}{2}]} g(\lambda, t)$

Note that  $\arg \min_{\lambda \geq 0} g(\lambda, t) = \frac{t}{\lambda^2}$  and  $\min_{\lambda \geq 0} g(\lambda, t) = -\frac{t^2}{2\lambda^2}$

so if  $0 \leq t < \frac{\lambda^2}{2}$  then  $g^*(t) = -\frac{t^2}{2\lambda^2}$

if  $t \geq \frac{\lambda^2}{2}$  then  $g(\cdot, t) \downarrow$  and

$$g^*(t) = g\left(\frac{1}{\lambda}, t\right) = -\frac{t}{2} + \frac{1}{2\lambda} \frac{\lambda^2}{\lambda} \leq -\frac{t}{2\lambda} \quad \square$$

#### IV] Martingale band method

- $\{X_k\}_{k=1}^n$  independent r.v.
- $f: \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f(x) := f(x_1, \dots, x_n)$
- $Y_0 = \mathbb{E}[f(x)]$  assume that  $\mathbb{E}[|f(x)|] < \infty$

$$Y_m = f(x)$$

$$Y_k = \mathbb{E}[f(x) | X_1, \dots, X_k], \quad k=1, \dots, m-1$$

Telescopic sum:

$$f(x) - \mathbb{E}[f(x)] = Y - Y_0 = \sum_{k=1}^n \underbrace{(Y_k - Y_{k-1})}_{D_k}$$

•  $\{Y_k\}_{k=1}^n$  is the Dobr martingale:

Martingale

$$\left\{ \begin{array}{l} 1/ \mathbb{E}|Y_k| = \mathbb{E}[\mathbb{E}[f(x)|X_1, \dots, X_k]] \\ \leq \mathbb{E}[|f(x)|] \quad \text{Jensen} \\ < \infty \end{array} \right.$$

$$\left\{ \begin{array}{l} 2/ \mathbb{E}[Y_{k+1} | X_1, \dots, X_k] = \mathbb{E}[\mathbb{E}[f(x)|X_1, \dots, X_{k+1}] | X_1, \dots, X_k] \\ = \mathbb{E}[f(x) | X_1, \dots, X_k] \quad \text{tower} \\ = Y_k \end{array} \right.$$

•  $\{D_k\}$  is a martingale difference sequence

$$1/ \mathbb{E}[|D_k|] < \infty$$

$$2/ \mathbb{E}[D_{k+1} | X_1, \dots, X_k] = 0$$

Theorem If  $\mathbb{E}[e^{\lambda D_k} | X_1, \dots, X_{k-1}] \leq e^{\frac{\lambda^2 V_k^2}{2}}$  as

for  $|\lambda| < \frac{1}{\alpha_k}$

Then a)  $\sum_{k=1}^m D_k$  is sub-exponential with parameters

$$\left( \sqrt{\sum_{k=1}^m V_k^2}, \underbrace{\max_{k=1, \dots, m} \alpha_k}_{\alpha_*} \right)$$

b) concentration inequality:

$$P[|f(x) - \mathbb{E} f(x)| > \epsilon] \leq \begin{cases} 2 e^{-\frac{\epsilon^2}{2 \sum_{k=1}^m V_k^2}} & \text{if } 0 \leq \epsilon \leq \frac{\sum_{k=1}^m V_k^2}{\alpha_*} \\ 2 e^{-\frac{\epsilon}{2 \alpha_*}} & \text{if } \epsilon > \frac{\sum_{k=1}^m V_k^2}{\alpha_*} \end{cases}$$

Proof:  $\mathbb{E}\left[e^{\lambda \left(\sum_{k=1}^m D_k\right)}\right]$

$$= \mathbb{E}\left[e^{\lambda \left(\sum_{k=1}^m D_k\right)} \mathbb{E}[e^{\lambda D_m} | X_1, \dots, X_{m-1}]\right]$$

$$\leq \mathbb{E}\left[e^{\lambda \left(\sum_{k=1}^m D_k\right)}\right] e^{\frac{\lambda^2 V_m^2}{2}} \quad \square$$

Cor: (Azuma-Hoeffding)

$D_k \in [a_k, b_k]$  as

$$\mathbb{P}[|f(x) - \mathbb{E} f(x)| \geq t] \leq 2 e^{-\frac{2t^2}{\sum_{k=1}^n (b_k - a_k)^2}}$$

Notation  $x, x' \in \mathbb{R}^n$ ,  $k \in \{1, \dots, n\}$

$$x^{(k)} = (x_1^{(k)}, \dots, x_n^{(k)}) \text{ and } x_j^{(k)} = \begin{cases} x_j & \text{if } j \neq k \\ x'_j & \text{if } j = k \end{cases}$$

We say that  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  enjoys the bounded difference inequality with parameters  $(L_1, \dots, L_n)$  if:

$$\forall k \in \{1, \dots, n\}, \quad |f(x) - f(x^{(k)})| \leq L_k, \quad \forall x, x' \in \mathbb{R}^n$$

Cor:  $\mathbb{P}[|f(x) - \mathbb{E} f(x)| \geq t] \leq 2 e^{-\frac{2t^2}{\sum_{k=1}^n L_k^2}}, \quad \forall t \geq 0$

Bounded difference inequality

$$\text{Proof: } D_k = \mathbb{E}[f(x) | x_1, \dots, x_k] - \mathbb{E}[f(x) | x_1, \dots, x_{k-1}]$$

$$\text{Set: } A_k = \inf_x \mathbb{E}[f(x) | x_1, \dots, x_{k-1}, x] - \mathbb{E}[f(x) | x_1, \dots, x_{k-1}] \quad \text{R.V.}$$

$$B_k = \sup_x \mathbb{E}[f(x) | x_1, \dots, x_{k-1}, x] - \mathbb{E}[f(x) | x_1, \dots, x_{k-1}]$$

$$\text{We have } B_k \geq D_k \geq A_k \text{ a.s.}$$

It suffices to prove that  $B_k - A_k \leq L_k$  a.s to conclude.

$$B_k - A_k = \sup_x \mathbb{E}[f(x) | x_1, \dots, x_{k-1}, x] - \mathbb{E}[f(x) | x_1, \dots, x_{k-1}] \\ - \inf_x \mathbb{E}[f(x) | x_1, \dots, x_{k-1}, x] - \mathbb{E}[f(x) | x_1, \dots, x_{k-1}]$$

$$\text{denote } E_{k+1} = E_{x_{k+1}, \dots, x_n} \text{ then}$$

$$B_k - A_k = \sup_x E_{k+1} [f(x_1, \dots, x_{k-1}, x, x_{k+1}, \dots, x_n)]$$

$$- \inf_x E_{k+1} [f(x_1, \dots, x_{k-1}, x, x_{k+1}, \dots, x_n)]$$

$$\leq \sup_{x,y} | E_{k+1} [f(x_1, \dots, x_{k-1}, x, x_{k+1}, \dots, x_n)] \\ - E_{k+1} [f(x_1, \dots, x_{k-1}, y, x_{k+1}, \dots, x_n)] |$$

$\leq L_k$  by B.D assumption  $\square$

## V] Lipschitz functions of Gaussian variables

$f: \mathbb{R}^m \rightarrow \mathbb{R}$   $L$ -lipschitz:

$$\forall x, y \in \mathbb{R}^m, |f(x) - f(y)| \leq L \|x - y\|_2$$

Theorem  $X \sim \mathcal{N}(0, \text{Id}_n)$   
 $f$   $L$ -lip

$Z = f(x) - \mathbb{E} f(x)$  is sub-gaussian  
with param. at most  $L$

$$\rightarrow \mathbb{P}[|Z| \geq \epsilon] \leq 2 e^{-\frac{\epsilon^2}{2L}}, \forall \epsilon \geq 0$$

Proof: We assume that  $f$  differentiable and

$$\|\nabla f(x)\|_2 \leq L$$

Lemma: For any convex function  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$  we have:

$$\mathbb{E}\left[\varphi\left[f(x) - \mathbb{E}f(x)\right]\right] \leq \mathbb{E}\left[\varphi\left(\frac{\pi}{2} \langle \nabla f(x), y \rangle\right)\right]$$

where  $X \perp\!\!\!\perp Y$  and  $X \sim Y \sim \mathcal{N}_n(0, \text{Id}_n)$

↳ proof later

Fix  $\lambda \in \mathbb{R}$  and consider  $\varphi(t) = e^{\lambda t}$ , it yields:

$$\begin{aligned} & \mathbb{E}_x \left[ \exp\left(\lambda(f(x) - \mathbb{E}f(x))\right) \right] \\ & \leq \mathbb{E}_{x,y} \left[ \exp\left(\frac{\lambda\pi}{2} \langle y, \nabla f(x) \rangle\right) \right] \\ & \stackrel{\text{exp. over } y}{=} \mathbb{E}_x \left[ \exp\left(\frac{\lambda^2\pi^2}{2} \|\nabla f(x)\|_2^2\right) \right] \\ & \stackrel{\text{Lip}}{\leq} e^{\frac{1}{8}\lambda^2\pi^2 L^2} \end{aligned}$$

→  $Z$  sub Gaussian with parameter at most  $\frac{\pi L}{2}$ .

→ tail bound follows.  $\square$

Proof of key lemma above:

- For  $\theta \in [0, \frac{\pi}{2}]$ , let

$$W(\theta) = X \sin \theta + Y \cos \theta$$

- By Jensen:

$$\mathbb{E}_x [\varphi(f(x) - \mathbb{E}_y f(y))] \leq \mathbb{E}_{x,y} [\varphi(f(x) - f(y))]$$

- Note that  $W(0) = Y$  and  $W\left(\frac{\pi}{2}\right) = X$  so

$$f(x) - f(y) = \int_0^{\frac{\pi}{2}} \frac{d}{d\theta} f(W(\theta)) d\theta$$

$$\stackrel{\text{chain rule}}{=} \int_0^{\frac{\pi}{2}} \langle \nabla f(W(\theta)), W'(\theta) \rangle d\theta$$

but  $| W'(\theta) = X \cos \theta - Y \sin \theta$

$$W'(\theta) \perp\!\!\!\perp W(\theta)$$

$$\begin{pmatrix} W(\theta) \\ W'(\theta) \end{pmatrix} \sim W(\theta, \text{Id}_{2m})$$

• Then :

$$\begin{aligned}
 \mathbb{E}_x \left[ \varphi(f(x) - \mathbb{E}_y f(y)) \right] &\leq \mathbb{E}_{x,y} \left[ \varphi \left( \int_0^{\frac{\pi}{2}} \langle \nabla f(w(\theta)), w'(\theta) \rangle d\theta \right) \right] \\
 &= \mathbb{E}_{x,y} \left[ \varphi \left( \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \frac{\pi}{2} \langle \nabla f(w(\theta)), w'(\theta) \rangle d\theta \right) \right] \\
 &\stackrel{\text{Fubini}}{\leq} \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \mathbb{E}_{x,y} \left[ \varphi \left( \frac{\pi}{2} \langle \nabla f(w(\theta)), w'(\theta) \rangle \right) \right] d\theta
 \end{aligned}$$

$$\text{Now } (w(\theta), w'(\theta)) \stackrel{d}{=} (\tilde{x}, \tilde{y}) \quad \text{whr } \begin{cases} \tilde{x} \perp \tilde{y} \\ \tilde{x} \sim \tilde{y} \sim U(0, 1) \end{cases}$$

$$= \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \mathbb{E} \left[ \varphi \left( \frac{\pi}{2} \langle \nabla f(\tilde{x}), \tilde{y} \rangle \right) \right] d\theta$$

$$= \mathbb{E} \left[ \varphi \left( \frac{\pi}{2} \langle \nabla f(\tilde{x}), \tilde{y} \rangle \right) \right]$$

□