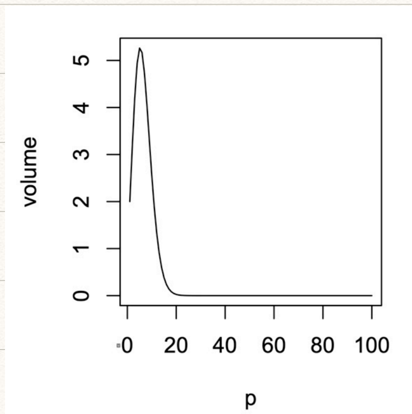
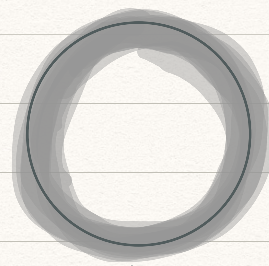
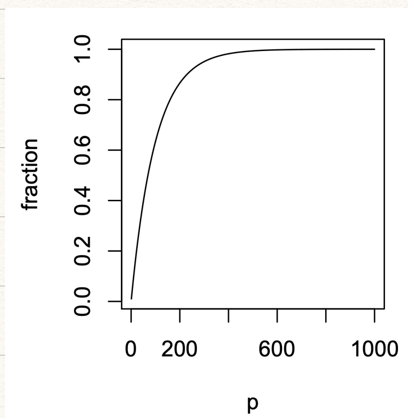
Stranger things

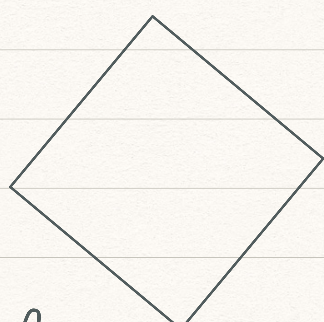


Volume of the Euclidean ball

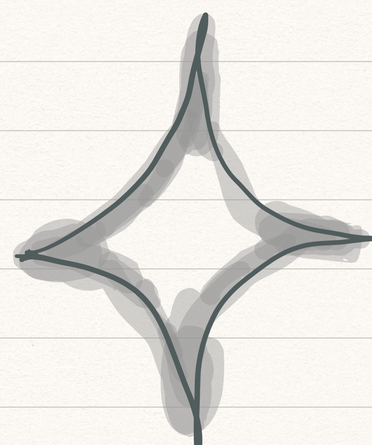




$$\text{fraction} = \frac{\mathrm{Vol}(0.99 \leq \|x\|_2 \leq 1)}{\mathrm{Vol}(\|x\|_2 \leq 1)}$$



l_1 ball



l_1 ball

Course 1 : Probability toolbox

0). Central limit theorem

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

$X_1, \dots, X_n$ iid s.t $\mathbb{V}\text{ar}[f(X_1)] < +\infty$

$$\frac{1}{n} \sum_{i=1}^{n} f(X_i) = \mathbb{E} f(X_1) + \left[\frac{\mathbb{V}\text{ar}[f(X_1)]}{n} \right]^{\frac{1}{2}} Z_n$$

with $Z_n \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0,1)$

Assume f $\underline{L\text{-Lipschitz}}$ $|f(x) - f(y)| \leq L|x-y|$

X_i finite variance σ^2

Then
$$\mathbb{V}\text{ar}[f(X_1)] = \frac{1}{2} \mathbb{E}\left[(f(X_1) - f(X_2))^2 \right]$$
$$\leq \frac{L^2}{2} \mathbb{E}\left[(X_1 - X_2)^2 \right]$$

$\underline{\mathbb{V}\text{ar}[f(X_1)] \leq L^2 \sigma^2}$

We deduce that for $x > 0$

$$\lim_{n \to +\infty} \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E} f(X_1) \geq \frac{L\sigma}{\sqrt{n}} x\right]$$

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

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

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

1/ $h'(x) = e^{-\frac{x^2}{2}}\left[\frac{1}{\sqrt{2\pi}} - x\right]$ which is ≥ 0 for $x \leq \frac{1}{\sqrt{2\pi}}$

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

1+2 : $h > 0$ for $x \in \left]0, \frac{1}{\sqrt{2\pi}}\right[$

3/ for $x > \frac{1}{\sqrt{2\pi}}$: $\frac{1}{\sqrt{2\pi}}\int_x^{\infty} e^{-\frac{t^2}{2}} dt \leq \int_x^{+\infty} t\, e^{-\frac{t^2}{2}} dt = e^{-\frac{x^2}{2}}$ $\square$

Goal Give a non asymptotic version of this result.

I] From Markov to Chernoff

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

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

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

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

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

Chernoff bound Moment Generating Function

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

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

Optimizing λ:

$$\boxed{\log \mathbb{P}[X-\mu \geq t] \leq \inf_{\lambda \in [0,b]} \left\{ \log[\varphi(\lambda)] - \lambda t \right\}}$$

Remark: Suppose $X \geq 0$, set:

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

then
$$\begin{aligned} e^{\lambda\delta} &= \sum_{k\geq 0} \frac{\lambda^k}{k!} \delta^k \\ &\leq C^{-1} \sum_{k\geq 0} \frac{\lambda^k}{k!} \mathbb{E}\left[|X|^k\right] \\ &= C^{-1} \mathbb{E}\, e^{\lambda X} \end{aligned}$$

hence
$$\boxed{\inf_{k\geq 0} \frac{\mathbb{E}\left[|X|^k\right]}{\delta^k} \leq \inf_{\lambda \geq 0} \frac{\mathbb{E}\, e^{\lambda X}}{e^{\lambda\delta}}}$$

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

II] Sub-Gaussian and Hoeffding

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

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

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

We uncover : $\boxed{\mathbb{P}\left[X - \mu \geq t\right] \leq e^{-\frac{t^2}{2\sigma^2}}}$

Def (Sub-Gaussian)

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

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

σ is refered to as the <u>sub-gaussian parameter</u>

<u>example</u> $\varepsilon \sim \mathrm{Unif}(\{-1, +1\})$ <u>Rademacher</u>

prove that $\sigma = 1$

$$\varphi_\varepsilon(\lambda) = \mathbb{E}[e^{\lambda\varepsilon}] \leq e^{\frac{\lambda^2}{2}}$$

<u>example</u> Bounded R.V. : $X \in [a,b]$

$$\varphi_X(\lambda) = \mathbb{E}_X[e^{\lambda(X - \mathbb{E}_{X'}[X'])}] \quad | X' \text{ ind. copy}$$

$$\overset{\text{Jensen}}{\leq} \mathbb{E}_{X,X'}[e^{\lambda(X-X')}] \left(= \int e^{\lambda(z-z')} f_{(X,X')}(z,z')\,dz\,dz' \right)$$

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

$$X - X' \overset{d}{=} X' - X \overset{d}{=} \varepsilon(X - X')$$

$$[(X-X') \sim \varepsilon(X-X')]$$

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

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

$\longrightarrow$ $\boxed{\sigma = b-a}$ one can prove: $\boxed{\sigma = \frac{b-a}{2}}$

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

Prop X_1, X_2 ind. sub-Gaussian with parameters σ_1, σ_2 then X_1+X_2 is sub-Gaussian with parameter $\sqrt{\sigma_1^2+\sigma_2^2}$.

proof: $\varphi_{X_1+X_2}(\lambda) = \mathbb{E}\, e^{\lambda(X_1+X_2-\mu_1-\mu_2)}$

$$\overset{\text{ind}}{=} \mathbb{E}\, e^{\lambda(X_1-\mu_1)}\, \mathbb{E}\, e^{\lambda(X_2-\mu_2)}$$

$$\leq e^{\frac{\lambda^2}{2}(\sigma_1^2+\sigma_2^2)} \qquad \square$$

Prop (Hoeffding bound)

If X_i, $i = 1, \dots, n$ are independent

X_i has mean μ_i and sub-Gaussian parameter σ_i

Then $\forall\, t \geq 0$,

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

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

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

$$\boxed{\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^n (X_i-\mu_i) \geq \frac{(b-a)}{\sqrt{n}}\sqrt{-\log \varepsilon}\right] \leq \varepsilon}$$

III) Sub-exponential

Def (Sub-exponential)

A r.v. X with mean μ is sub-exponential if there are non-negative parameters (ν, α) s.t.

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

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

One can check that:

- $\mu = \mathbb{E}X = 1$
- $$\varphi_X(\lambda) = \mathbb{E}\left[e^{\lambda(X-1)}\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\lambda(z^2-1)} e^{-\frac{z^2}{2}} dt = \frac{e^{-\lambda}}{\sqrt{1-2\lambda}}$$

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

$\rightarrow$ X is not Sub-Gaussian

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

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

Prop (Sub-exponential tail bound)

X sub-exponential (ν, α)

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

Proof: Wlog, $\mu = 0$

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

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

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

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

if $t \geq \frac{\nu^2}{\alpha}$ then $g(\cdot, t) \searrow$ and

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

IV) Martingale bound method

- $\{X_k\}_{k=1}^{n}$ independent r.v.
- $f : \mathbb{R}^n \to \mathbb{R}$, $f(X) := f(X_1, \dots, X_n)$
- $Y_0 = \mathbb{E}[f(X)]$ assume that $\mathbb{E}[|f(X)|] < \infty$

$Y_n = f(X)$

$Y_k = \mathbb{E}[f(X) \mid X_1, \dots, X_k]$, $k = 1, \dots, n-1$

Telescopic sum:

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

- $\{Y_k\}_{k=1}^{n}$ is the <u>Doob martingale</u>:

Martingale
$$\begin{cases} 1/\ \mathbb{E}|Y_k| = \mathbb{E}\left[\,|\mathbb{E}[f(x)\,|\,X_1,\dots,X_k]|\,\right] \\ \qquad \le \mathbb{E}[\,|f(x)|\,] \qquad \underline{\text{Jensen}} \\ \qquad < \infty \\ 2/\ \mathbb{E}[Y_{k+1}\,|\,X_1,\dots,X_k] = \mathbb{E}\left[\mathbb{E}[f(x)\,|\,X_1,\dots,X_{k+1}]\,|\,X_1,\dots,X_k\right] \\ \qquad = \mathbb{E}[f(x)\,|\,X_1,\dots,X_k] \qquad \underline{\text{tower}} \\ \qquad = Y_k \end{cases}$$

- $\{D_k\}$ is a <u>martingale difference sequence</u>

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

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

Theorem If $\mathbb{E}\left[e^{\lambda D_k} \mid X_1, \dots, X_{k-1}\right] \leq e^{\lambda^2 \frac{\nu_k^2}{2}}$ a.s

for $|\lambda| < 1/\alpha_k$

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

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

b) concentration inequality:

$$\mathbb{P}\left[|f(X) - \mathbb{E} f(X)| > t \right] \leq \begin{cases} 2 e^{-\frac{t^2}{2\sum_{k=1}^{n} \nu_k^2}} & \text{if } 0 \leq t \leq \frac{\sum_{k=1}^{n} \nu_k^2}{\alpha_*} \\ 2 e^{-\frac{t}{2\alpha_*}} & \text{if } t > \frac{\sum_{k=1}^{n} \nu_k^2}{\alpha_*} \end{cases}$$

Proof: $\mathbb{E}\left[e^{\lambda \left(\sum_{k=1}^{n} D_k \right)} \right]$, $\quad |\lambda| \leq \frac{1}{\alpha_*}$

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

$$\leq \mathbb{E}\left[e^{\lambda \left(\sum_{k=1}^{n-1} D_k \right)} \right] e^{\lambda^2 \nu_n^2 / 2} \qquad \square$$

Cor: (Azuma-Hoeffding)

$D_k \in [a_k, b_k]$ as

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

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

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

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

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

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

Bounded difference inequality

Proof: $D_k = \mathbb{E}[f(x) \mid X_1, \dots, X_k] - \mathbb{E}[f(x) \mid X_1, \dots, X_{k-1}]$

Set: $A_k = \inf_x \mathbb{E}[f(x) \mid X_1, \dots, X_{k-1}, x] - \mathbb{E}[f(x) \mid X_1, \dots, X_{k-1}]$

$B_k = \sup_x \mathbb{E}[f(x) \mid X_1, \dots, X_{k-1}, x] - \mathbb{E}[f(x) \mid X_1, \dots, X_{k-1}]$

R.V.

We have $B_k \geq D_k \geq A_k$ a.s.

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

$$B_k - A_k = \sup_x \mathbb{E}[f(x) \mid X_1, \dots, X_{k-1}, x] - \cancel{\mathbb{E}[f(x) \mid X_1, \dots, X_{k-1}]}$$
$$- \inf_x \mathbb{E}[f(x) \mid X_1, \dots, X_{k-1}, x] - \cancel{\mathbb{E}[f(x) \mid X_1, \dots, X_{k-1}]}$$

denote $\mathbb{E}_{k+1} = \mathbb{E}_{X_{k+1}, \dots, X_n}$ then

$$B_k - A_k = \sup_x \mathbb{E}_{k+1}[f(X_1, \dots, X_{k-1}, x, X_{k+1}, \dots, X_n)]$$
$$- \inf_y \mathbb{E}_{k+1}[f(X_1, \dots, X_{k-1}, y, X_{k+1}, \dots, X_n)]$$

$(-\inf h(y) = \sup(-h(y)))$

$$\leq \sup_{x,y} \Big| \mathbb{E}_{k+1}[f(X_1, \dots, X_{k-1}, x, X_{k+1}, \dots, X_n)] - \mathbb{E}_{k+1}[f(X_1, \dots, X_{k-1}, y, X_{k+1}, \dots, X_n)] \Big|$$

$$\leq \sup_{x,y} \mathbb{E}_{k+1}\Big[\underbrace{|f(*, x, *) - f(*, y, *)|}_{\leq L_k}\Big]$$

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

V] Lipschitz functions of Gaussian variables

$f : \mathbb{R}^n \to \mathbb{R}$ L-lipschitz :

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

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

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

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

Proof: We assume that f differentiable and

$$\boxed{\|\nabla f(x)\|_2 \leq L}$$

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

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

when $X \perp\!\!\!\perp Y$ and $X \sim Y \sim \mathcal{N}_n(0, I_{d_n})$

$\hookrightarrow$ proof later

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

$$\mathbb{E}_X\left[\exp\left(\lambda\left(f(X) - \mathbb{E}_X f(X)\right)\right)\right]$$

$$\leq \underbrace{\mathbb{E}_{X,Y}\left[\exp\left(\frac{\lambda\pi}{2} \langle Y, \nabla f(X) \rangle\right)\right]}_{\text{MGF of } Y \sim \mathcal{N}(0, I_d)}$$

exp. over Y

$$= \mathbb{E}_X\left[\exp\left(\frac{\lambda^2\pi^2}{8} \|\nabla f(X)\|_2^2\right)\right]$$

$$\overset{\text{Lip}}{\leq} e^{\frac{1}{8}\lambda^2\pi^2 L^2}$$

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

$\to$ tail bound follows. $\square$

Proof of key lemma above:

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

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

- By Jensen:

$$\mathbb{E}_X \left[\varphi\left(f(X) - \mathbb{E}_Y f(Y) \right) \right] \leq \mathbb{E}_{X,Y} \left[\varphi\left(f(X) - f(Y) \right) \right]$$

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

$$f(X) - f(Y) = \int_0^{\frac{\pi}{2}} \frac{d}{d\theta} f(W(\theta)) \, d\theta$$

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

but

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

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

$\mathrm{Cov}(W'(\theta); W(\theta))$
$= \mathrm{Cov}(X \sin\theta + Y\cos\theta ; X\cos\theta - Y \sin\theta)$
$= \sin\theta\cos\theta \, Id - \sin\theta\cos\theta \, Id = 0$

$$\begin{pmatrix} W(\theta) \\ W'(\theta) \end{pmatrix} \sim \mathcal{N}(0, Id_{2n})$$

· Then :

$$\mathbb{E}_X\left[\varphi\left(f(X)-\mathbb{E}_Y f(Y)\right)\right] \le \mathbb{E}_{X,Y}\left[\varphi\left(\int_0^{\frac{\pi}{2}} \langle \nabla f(W(\theta)), W'(\theta)\rangle \, d\theta\right)\right]$$

$$= \mathbb{E}_{X,Y}\left[\varphi\left(\frac{2}{\pi}\int_0^{\frac{\pi}{2}} \frac{\pi}{2}\langle \nabla f(W(\theta)), W'(\theta)\rangle \, d\theta\right)\right]$$

Jensen

Fubini

$$\le \frac{2}{\pi}\int_0^{\frac{\pi}{2}} \mathbb{E}_{X,Y}\left[\varphi\left(\frac{\pi}{2}\langle \nabla f(W(\theta)), W'(\theta)\rangle\right)\right] d\theta$$

Now $\left(W(\theta), W'(\theta)\right) \overset{d}{=} (\tilde{X}, \tilde{Y})$ when $\tilde{X} \perp\!\!\!\perp \tilde{Y}$, $\tilde{X} \sim \tilde{Y} \sim \mathcal{N}(0, I_d)$

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

$$= \mathbb{E}\left[\varphi\left(\frac{\pi}{2}\langle \nabla f(\tilde{X}), \tilde{Y}\rangle\right)\right] \qquad \square$$