Info Retrieval Assignment 1

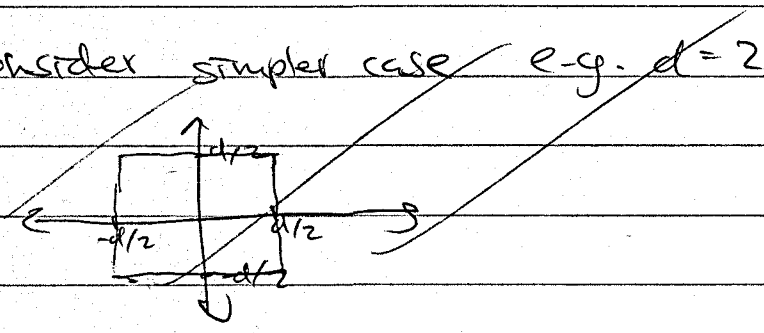
Curse of Dimensionality

Task 1: $C_d = \{ \vec{x} \in \mathbb{R}^p \mid \|x\|_\infty \leq \frac{d}{2} \}$ (p-dimensional hypercube) w/ length d

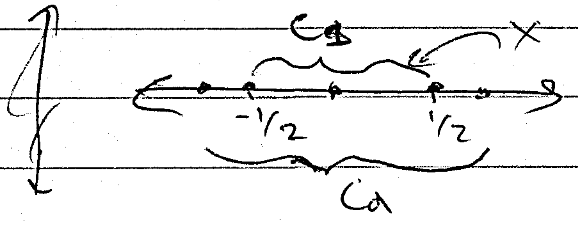
denotes max value of occurring elements in vector.

a) Assume X ~ Uniformly distributed inside C,
- Determine d in dependence of p and $q \in [0,1]$ so $Pr(X \in C_d) = q$ holds.

Consider simpler case e.g. d = 2



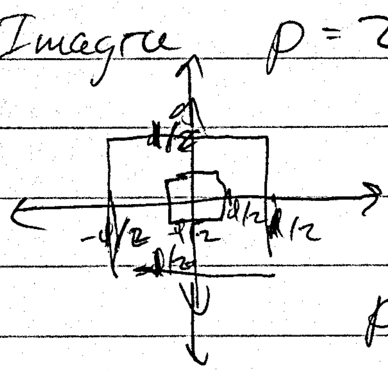
What is C_d for p=1? Imagine in p=1



$$Pr(X \in C_d) = \begin{cases} 1, & d > 1 \\ d, & d < 1 \end{cases} = q$$

$p=1$: $d = q$, $d < 1$

Imagine p = 2



$$Pr(X \in C_d) = \begin{cases} 1, & d > 1 \\ d^2, & d < 1 \end{cases} = q$$

$p = 2$: $d^2 = q$, $d < 1$

$d = \sqrt{q}$

Generalise:

$p=1$: $d = q$, $d < 1$

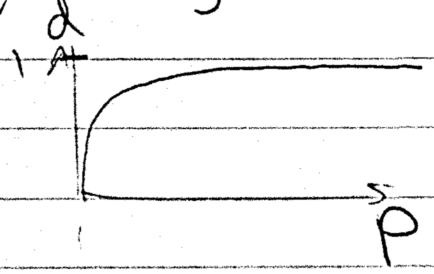
$p=2$: $d = \sqrt{q}$, $d < 1$

then ~~$d = p\sqrt{q}$~~, $d < 1$

$d = \sqrt[p]{q} = q^{1/p}$

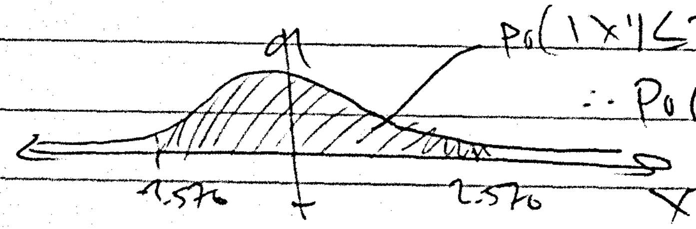
What does this mean?

- as dimensionality increases, probability ~~of being inside~~ samples being inside ~~(d decreases~~ the d required to sustain the probability increases

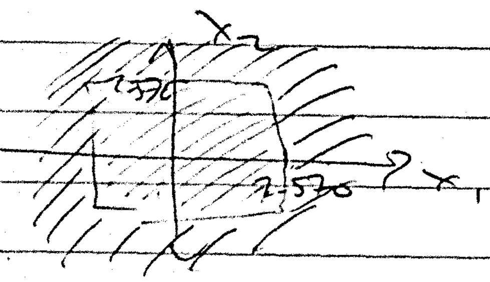


$(q<1)$ $q \in (0,1)$

b) Visualise (1)



What about $Pr(\|X^2\|_\infty > 2.570)$?



$Pr(\|X^2\|_\infty > 2.576) = Pr(|X_1| > 2.576 \cup |X|_2 > 2.576)$

~~$= P(|X_1| > 2.576$~~

$= 1 - Pr(\|X^2\|_\infty < 2.570)$

$= 1 - Pr(|X_1| < 2.576 \wedge |X_2| < 2.576)$

$= 1 - Pr(|X_1| < 2.576)\, Pr(|X_2| < 2.576)$

$= 1 - \cancel{(0.01}(0.99)(0.99)$

$= 0.9801$

Generalise: $Pr(\|X^p\|_\infty > 2.576) = 1 - 0.99^p$

$p = 1$ $\quad Pr(\;) = 0.01$

$= 2$ $\quad = \cancel{0.980} + 0.0199$

$= 3$ $\quad = 0.0297$

$= 500$ $\quad = 0.993$

As dimensionality increases, probability of lying outside hypercube of fixed of increases.

Statistical Decision Making

a) Figure 1 :

$X=x$			
2	0.4	0.14	0.05
1	0.02	0.26	0.13
	1	2	3 $Y=y$

Is this probability table?

$Sum = 0.4 + 0.14 + 0.05 + 0.02 + 0.26 + 0.13$
$= \underline{\underline{1}}$

Yes: sums to 1

b) (i) Find $E_{Y|X=2}[Y]$
(ii) $Pr(X=1 \mid Y=3)$

(i) $p_X(x) = \begin{cases} 0.59, & x=2 \\ 0.41, & x=1 \end{cases}$

$p_{Y|X=2}(y) = \cancel{\begin{cases} 0.14/0.59 = \\ 0.26/0.59 = 0.68 \end{cases}}$

$= \begin{cases} 0.4/0.59 = 0.678, & y=1 \\ 0.14/0.59 = 0.237, & y=2 \\ 0.05/0.59 = 0.085, & y=3 \\ 0, & \text{else} \end{cases}$

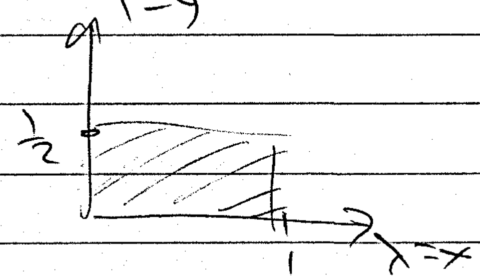
$E_{Y|X=2}[Y] = 1(0.678) + 2(0.237) + 3(0.085)$
$= \underline{\underline{1.407}}$

(ii) $P_r(X=1 \mid Y=3) = \dfrac{P(X=1 \wedge Y=3)}{Pr(Y=3)} = \dfrac{0.13}{0.05+0.13} = 0.722$

c) Is $p(x,y) = \begin{cases} 1, & 0 \le x \le 1, \quad 0 \le y \le 1/2 \\ 0, & \text{else} \end{cases}$

a joint density function?

$\cancel{\iint p(x,y)}$ Visualise



$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\,dx\,dy = \int_0^{1/2}\int_0^1 1\,dx\,dy$$

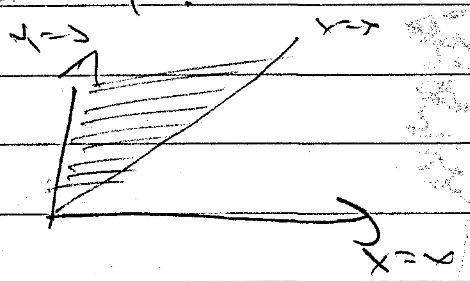
$$= \int_0^{1/2} 1\,dy$$

$$= 1/2 \neq 1$$

$\therefore$ not a joint density function

d) $p(x,y) = \begin{cases} 2e^{-(x+y)}, & 0 \le x < y,\ 0 \le y \\ 0, & \text{otherwise} \end{cases}$

Marginal density for X and Y?



$\cancel{p_X(x) = \int_{\infty} p(x,y)\,dy}$

$$p_X(x) = \int_{x}^{\infty}\int_0^{u} p(u,y)\,dy\,du$$

$$= \int_x^{\infty}\int_0^{u} p(u,y)\,dy\,du$$

$$= \int_x^{\infty}\int_0^{u} e^{-(u+y)}\,dy\,du$$

$$= \int_x^{\infty}\int_0^{u} e^{-u}e^{-y}\,dy\,du$$

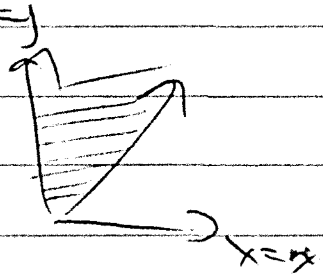
$$= -\int_x^{\infty} e^{-u}\left[e^{-y}\right]_{y=0}^{y=u} du = -\int_x^{\infty} e^{-u}(e^{-u}-1)\,du$$

$$= -\int_x^{\infty} e^{-2u}\,du + \int_x^{\infty} e^{-u}\,du$$

~~Then for y~~

$$p_Y(y) = \int_0^y 2e^{-(x+y)}\,dx$$
$$= 2e^{-y}\int_0^y e^{-x}\,dx$$
$$= 2e^{-y}\left[-e^{-x}\right]_0^y$$
$$= 2e^{-y}\left[-e^{-y}+e^{0}\right]$$
$$= 2e^{-y}\left[1-e^{-y}\right]$$
$$= 2e^{-y} - 2e^{-2y}$$

This is right



but $p_X(x)$ is wrong

Integrate dy ~~over~~ from x to ∞

$$p_X(x) = \int_x^\infty \cancel{p_Y(y)}\, p(x,y)\,dy$$
$$= \int_x^\infty 2e^{-(x+y)}\,dy$$
$$= 2e^{-x}\int_x^\infty e^{-y}\,dy$$
$$= 2e^{-x}\left[-e^{-y}\right]_{y=x}^{y=\infty}$$
$$= 2e^{-x}\left[-e^{-\infty}+e^{-x}\right]$$
$$= 2e^{-2x}$$

Check: $\int_0^\infty 2e^{-2x}\,dx = \left[-e^{-x}\right]_{x=0}^{x=\infty}$
$$= -e^{-\infty}+e^{0}$$
$$= 1$$