Name: Enbo Yu, ID: 113094714, Date: 08 SEP

1. (a) $f(x) = \|x\|_2^2$ is <u>not</u> a linear function.

if this function is linear, so for $\forall \alpha, \beta \in R$, $x, y \in R^n$

we should have $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$

now we let $d = 2$, and we can let $x = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $y = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

and let $\alpha = 1$, $\beta = 1$. so:

$$\alpha x + \beta y = 1 \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} + 1 \cdot \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$f(\alpha x + \beta y) = (\sqrt{0^2+0^2})^2 = 0$$

$$\alpha f(x) = 1 \cdot (\sqrt{0^2+1^2})^2 = 1$$

$$\beta f(y) = 1 \cdot (\sqrt{0^2+(-1)^2})^2 = 1$$

so $f(\alpha x + \beta y) \neq \alpha f(x) + \beta f(y)$.

<u>so $f(x) = \|x\|_2^2$ is not linear.</u>

(b) $f(x) = \|x\|_1$ is <u>not</u> a linear function.

as proved in section (a). we can let $d = 2$.

let $x = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, let $y = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$,

let $\alpha = 1$, $\beta = 1$.

$$\alpha x + \beta y = 1 \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} + 1 \cdot \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

so $f(\alpha x + \beta y) = f(\begin{bmatrix} 0 \\ 0 \end{bmatrix}) = 0 + 0 = 0$

$$\alpha \cdot f(x) = 1 \cdot f(\begin{bmatrix} 0 \\ 1 \end{bmatrix}) = 1 \cdot (0 + 1) = 1$$

$$\beta \cdot f(y) = 1 \cdot f(\begin{bmatrix} 0 \\ -1 \end{bmatrix}) = 1 \cdot (0 + |-1|) = 1$$

so $f(\alpha x + \beta y) \neq \alpha f(x) + \beta f(y)$.

<u>so $f(x) = \|x\|_1$ is not linear.</u>

(C) if $Q=0$ and $r=0$, $f(x)$ is a linear function.

if $Q\neq 0$ or $r\neq 0$, $f(x)$ is not a linear function.

(i) when $Q=0$ and $r=0$, we get

$f(x)= p^T x$, so for $\forall \alpha,\beta \in R$, $x,y\in R^n$

$$f(\alpha x+\beta y) = p^T\alpha x + p^T\beta y$$
$$= \alpha\cdot p^T x + \beta\cdot p^T y$$
$$= \alpha f(x) + \beta f(y).$$

so $f(x)$ is linear.

(ii) when $Q\neq 0$ or $r\neq 0$;

we can let $d=2$, let $Q=\begin{bmatrix}1&0\\0&1\end{bmatrix}$, $P=\begin{bmatrix}1\\1\end{bmatrix}$, $r=1$.

let $\alpha=\beta=1$,

$x=\begin{bmatrix}0\\1\end{bmatrix}$, $y=\begin{bmatrix}0\\-1\end{bmatrix}$

$$f(x) = \frac{1}{2}x^TQx + p^Tx + r$$
$$= \frac{1}{2}x^T\cdot 1\cdot x + x[1] + x[2] + 1$$
$$= \frac{1}{2}(x[1]^2 + x[2]^2) + x[1] + x[2] + 1$$

$$\alpha x + \beta y = 1\cdot\begin{bmatrix}0\\1\end{bmatrix} + 1\cdot\begin{bmatrix}0\\-1\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}$$

$$f(\alpha x+\beta y) = 1$$
$$f(\alpha x) = 1\cdot\left[\frac{1}{2}(0^2+1^2) + 0 + 1 + 1\right] = \frac{5}{2}$$
$$f(\beta y) = 1\left[\frac{1}{2}(0^2+1^2) + 0 - 1 + 1\right] = \frac{1}{2}$$

so $f(\alpha x+\beta y)\neq f(\alpha x)+f(\beta y)$

so $f(x)$ is not linear.

(d) $f(x) = c^Tx + b^TAx$ is linear:

for $\forall\alpha,\beta\in R$, $x,y\in R^n$

$$f(\alpha x+\beta y) = c^T(\alpha x+\beta y) + b^TA(\alpha x+\beta y)$$

$$= C^T\alpha x + b^T A \alpha x + C^T \beta y + b^T \beta y$$

$$= \alpha(C^T x + b^T x) + \beta(C^T y + b^T y)$$

$$= \alpha f(x) + \beta f(y).$$

so <u>$f(x)$ is linear</u>.

2. if $f(x)$ is norm, the $f(x)$ should satisfy the following:

① $\|\beta x\| = |\beta| \cdot \|x\|$ Homogeneity

② $\|x+y\| \le \|x\| + \|y\|$ Δ - Inequality

③ $\|x\| \ge 0$, $\forall x$ Non - negativity

④ $\|x\| = 0 \Longleftrightarrow x = 0$

(a) Not norm:

we can let $d = 2$. let $x = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$

then $f(x) = \sum_k x[k] = -1 + (-1) = -2 < 0$

so the direct sum $f(x)$ can not satisty

" ③ $\|x\| \ge 0$, $\forall x$ Non - negativity "

So <u>it is not norm</u>.

(b) Not norm:

we can let $d = 2$. so

$$f(x+y) = \left(\sum_{k=1}^{d=2} \sqrt{|x[k] + y[k]|}\right)^2$$

$$= \left(\sqrt{|x[1]+y[1]|} + \sqrt{|x[2]+y[2]|}\right)^2$$

$$= |x[1]+y[1]| + |x[2]+y[2]| + 2\sqrt{|x[1]+y[1]|} \cdot \sqrt{|x[2]+y[2]|}$$

$$f(x) = \left(\sum_{k=1}^{d=2} \sqrt{|x[k]|}\right)^2$$

$$= \left(\sqrt{|x[1]|} + \sqrt{|x[2]|}\right)^2$$

$= |x[1]| + |x[2]| + 2\sqrt{|x[1]|\cdot|x[2]|}$

$f(y) = |y[1]| + |y[2]| + 2\sqrt{|y[1]|\cdot|y[2]|}$

so, for $x = \begin{bmatrix}1\\0\end{bmatrix}$, $y = \begin{bmatrix}0\\1\end{bmatrix}$, $x+y = \begin{bmatrix}1\\1\end{bmatrix}$

we can get $f(x+y) = 4$

$f(x) + f(y) = 1+1 = 2$, so $f(x+y) > f(x) + f(y)$

" ② $\|x+y\| \le \|x\| + \|y\|$ Δ-inequality" can not be satisfied, so <u>it is not norm.</u>

(C) not norm.

let $d = 2$, $x = 0$, then

$f(x) = -\sum_{i=1}^{d=2} (x[i] + \frac{1}{2}) \log_2 (x[i] + \frac{1}{2})$

$= -(\frac{1}{2}) \cdot \log_2 (\frac{1}{2}) = -\frac{1}{2} \cdot (-1) = \frac{1}{2} \neq 0$

so " ④ $\|x\| = 0 \iff x = 0$ " can not be satisfied.

so <u>it is not norm.</u>

(d) norm.

(i) $f(\beta x) = \sqrt{\sum_{k=1}^{d} \frac{|\beta x[k]|^2}{k}}$

$= \sqrt{\beta^2 \sum_{k=1}^{d} \frac{|x[k]|^2}{k}}$

$= \beta \cdot \sqrt{\sum_{k=1}^{d} \frac{|x[k]|^2}{k}}$

$= \beta \cdot f(x)$

so it can satisfy " ① - Homogeneity " ;

(ii) for $\forall x$, obviously $f(x) \ge 0$,

so it can satisfy " ③ - Non-negativity ";

(iii) if $x = 0$. then

$$f(x) = \sqrt{\sum_{k=1}^{d} \frac{0}{k}} = 0$$

so $\|x\| = 0$

If $\|x\| = 0$, then we get

$$\sqrt{\frac{x[1]^2}{1} + \frac{x[2]^2}{2} + \cdots + \frac{x[d]^2}{d}} = 0$$

obviously, every item is ≥ 0.

so $x[1]^2 = x[2]^2 = \cdots\cdots = x[d]^2 = 0$

so $x = 0$.

so it can satisfy "④ $\|x\| = 0 \Leftrightarrow x = 0$",

(iv) for any x, y.

$$\|x+y\| = \sqrt{\sum_{k=1}^{d} \frac{(x[k]+y[k])^2}{k}}$$

$$\|x+y\|^2 = \sum_{k=1}^{d} \frac{(x[k]+y[k])^2}{k}$$

$$= \sum_{k=1}^{d} \frac{x[k]^2 + 2x[k]y[k] + y[k]^2}{k}$$

$$= \cancel{\sum_{k=1}^{d} \frac{x[k]^2}{k}} + \cancel{\sum_{k=1}^{d} \frac{y[k]^2}{k}} + \cancel{2} \cdot \sum_{k=1}^{d} \frac{x[k]y[k]}{k}$$

for $\|x\| + \|y\|$:

$$(\|x\| + \|y\|)^2 = \left(\sqrt{\sum_{k=1}^{d} \frac{x[k]^2}{k}} + \sqrt{\sum_{k=1}^{d} \frac{y[k]^2}{k}}\right)^2$$

$$= \left(\sqrt{\sum_{k=1}^{d} \frac{x[k]^2}{k}}\right)^2 + 2 \cdot \sqrt{\sum_{k=1}^{d} \frac{x[k]^2}{k}} \cdot \sqrt{\sum_{k=1}^{d} \frac{y[k]^2}{k}} + \left(\sqrt{\sum_{k=1}^{d} \frac{y[k]^2}{k}}\right)^2$$

$$= \sum_{k=1}^{d} \frac{x[k]^2}{k} + \sum_{k=1}^{d} \frac{y[k]^2}{k} + 2\sqrt{\sum_{k=1}^{d} \frac{x[k]^2}{k}} \cdot \sqrt{\sum_{k=1}^{d} \frac{y[k]^2}{k}}$$

$$= \cancel{\sum_{k=1}^{d} \frac{x[k]^2}{k}} + \cancel{\sum_{k=1}^{d} \frac{y[k]^2}{k}} + \cancel{2} \cdot \sqrt{\sum_{k=1}^{d} \frac{x[k]^2}{k}} \cdot \sqrt{\sum_{k=1}^{d} \frac{y[k]^2}{k}}$$

so the difference of above is

$$\sqrt{\sum_{k=1}^{d}\frac{x[k]^2}{k}}\cdot\sqrt{\sum_{k=1}^{d}\frac{y[k]^2}{k}} - \sum_{k=1}^{d}\frac{x[k]y[k]}{k},$$

let vector $a=\begin{bmatrix}\frac{x[1]}{\sqrt{1}}\\ \frac{x[2]}{\sqrt{2}}\\ \vdots\\ \frac{x[d]}{\sqrt{k}}\end{bmatrix}$, $b=\begin{bmatrix}\frac{y[1]}{\sqrt{1}}\\ \frac{y[2]}{\sqrt{2}}\\ \vdots\\ \frac{y[k]}{\sqrt{k}}\end{bmatrix}$,

by Cauchy Schuartz inequality, we know

$$\|a\|_2\|b\|_2 \geq a^Tb$$

so $\sqrt{\sum_{k=1}^{d}\frac{x[k]^2}{k}}\cdot\sqrt{\sum_{k=1}^{d}\frac{y[k]^2}{k}} \geq \sum_{k=1}^{d}\frac{x[k]y[k]}{k}$.

so that means

$$(\|x\|+\|y\|)^2 - \|x+y\|^2 \geq 0$$

so it can satisfy " ③ $\|x+y\| \leq \|x\|+\|y\|$
Δ Inequality"

based on above 4 properties proving.

it is norm.

3. (a) if A & B are Independent, we have

$$P_{A,B}(a.b) = P_A(a)\cdot P_B(b).$$

for b= hat, a=red, $P_{A,B}(a.b) = 0.075$

$P_A(a)\cdot P_B(b) = 0.25\cdot 0.3 = 0.075 = P_{A,B}(a.b)$

for b= hat, a= blue, $P_{A,B}(a.b) = 0.075$

$P_A(a)\cdot P_B(b) = 0.25\cdot 0.3 = 0.075 = P_{A,B}(a,b)$

for b= hat, a= green. $P_{A,B}(a.b) = 0.15$

$P_A(a)\cdot P_B(b) = 0.5\cdot 0.3 = 0.15 = P_{A,B}(a.b)$

for b= T-shirt. a= red. $P_{A,B} = (a.b) = 0.075$

$P_A(a)\cdot P_B(b) = 0.25\cdot 0.3 = 0.075 = P_{A,B}(a.b)$

for b= shoes, a= green, $P_{A,B}(a.b) = 0.1$

$P_A(a) \cdot P_B(b) = 0.5 \cdot 0.2 = 0.1 = P_{A,B}(a,b)$

so for all matches of A and B, we can always get $P_A(a) \cdot P_B(b) = P_{A,B}(a,b)$

so A and B are Independent.

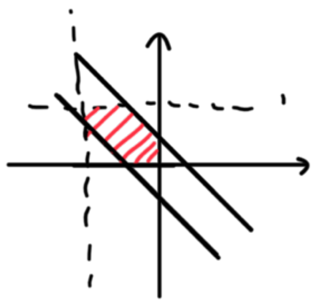
(b) $F_A(a) = \int_{-1}^{a} f_A(a) = x|_a - x|_{-1} = a+1 \quad (-1 \le a \le 0)$

$F_B(b) = \int_0^b f_B(b) = x|_b - x|_0 = b$

so $F_A(a) = \begin{cases} 1, & a > 0 \\ a+1, & -1 \le a \le 0 \\ 0, & a < -1 \end{cases}$ $\quad f_B(b) = \begin{cases} 1, & b > 1 \\ b, & 0 \le b \le 1 \\ 0, & b < 0 \end{cases}$

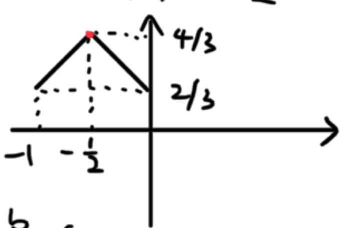
so $F_A(a) \cdot F_B(b) = \begin{cases} 1, & a > 0 \text{ and } b > 1 \\ 0, & a < -1 \text{ or } b < 0 \\ (a+1)b, & \text{else} \end{cases}$

$\int_{-\infty}^{+\infty} f_{AB}(a,b)\, db =$



$\begin{cases} \int_{-a-1/2}^{1} \frac{4}{3}\, db = \frac{4}{3}\left(\frac{3}{2} + a\right), & -1 < a < -\frac{1}{2} \\ \int_{0}^{-a+1/2} \frac{4}{3}\, db = \frac{4}{3}\left(\frac{1}{2} - a\right), & -\frac{1}{2} < a < 0 \end{cases}$

so f_A should be



$F_{A,B}(a,b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f_{A,B}(a,b)\, da\, db$

$= \int_{-\infty}^{a} \int_{-\infty}^{b} 4/3 \, da\, db$

obviously $F_{A,B}(a,b) \neq F_A(a) \cdot F_B(b)$

so A and B are not Independent.

(c) $B = A \cdot C = \begin{cases} 1 \\ -1 \end{cases}$

$P_B(b=1) = P_a(a=1) \cdot P_c(C=1) + P_a(a=-1) \cdot P_c(C=-1)$

$= 0.5 \cdot 0.9 + 0.5 \cdot 0.1$

$= 0.5$

$P_B(b=-1) = P_a(a=1) \cdot P_c(C=-1) + P_a(a=-1) \cdot P_c(C=1)$

$= 0.5 \cdot 0.1 + 0.5 \cdot 0.9$

$= 0.5$

$P_{A,B}(a=1, b=1) = \frac{1}{4} = P_B(b=1) \cdot P_A(a=1)$

$P_{A,B}(a=1, b=-1) = \frac{1}{4} = P_B(b=-1) \cdot P_A(a=1)$

$P_{A,B}(a=-1, b=1) = \frac{1}{4} = P_B(b=1) \cdot P_A(a=-1)$

$P_{A,B}(a=-1, b=-1) = \frac{1}{4} = P_B(b=-1) \cdot P_A(a=-1)$

so for all condition we have

$P_{A,B} = P_A \cdot P_B$

so A and B are independent.

(d) $\Sigma_{11} = E[(A - E[A])^2] = E[(A-0)^2] = E[A^2] = 1$

$\Sigma_{12} = \Sigma_{21} = E[(A - E[A])(B - E[B])]$

$= E[(A-0)(B-1)]$

$= E[A(B-1)] = -1$

$\Sigma_{22} = E[(B - E[B])^2] = E[(B-1)^2] = \frac{1}{2}$

so $\Sigma = \begin{bmatrix} 1 & -1 \\ -1 & \frac{1}{2} \end{bmatrix}$, not diagonal

so A and B are not independent.

4. (a) $P(\text{A wear}) = P(\text{A wear} \mid \text{Mon or Thu}) \cdot P(\text{Mon or Thu})$

$+ P(\text{A wear} \mid \text{other 5 days}) \cdot P(\text{other 5 days})$

$= 1 \cdot \frac{2}{7} + \frac{1}{2} \cdot \frac{5}{7} = \frac{9}{14}$

$P(\text{Z wear}) = \frac{4}{7}$

so the probability of either Alexa or Zuckie wearing a sock

$= P(\text{A wear}) + P(\text{Z wear})$

$- P(\text{A wear and Z wear})$

$= \frac{9}{14} + \frac{4}{7} - \left(\frac{9}{14} \cdot \frac{4}{7}\right) = \frac{63}{98} + \frac{56}{98} - \frac{36}{98}$

$= \frac{83}{98} = 0.84693\ldots\ldots \approx \underline{0.847}$

(b) Alexa, Siri = girl

Googs, Zuckie = boy.

$P(\text{A wear}) = \frac{9}{14} \rightarrow$ girl

$P(\text{S wear}) = 1 \rightarrow$ girl

$P(\text{G wear}) = 0 \rightarrow$ boy

$P(\text{Z wear}) = \frac{4}{7}$ — boy.

$P(\text{A wear} \mid \text{a girl wear})$

$$= \frac{P(\text{A wear, a girl wear})}{P(\text{a girl wear})}$$

$$= \frac{P(\text{A wear, a girl wear})}{\begin{aligned} &P(\text{a girl wear} \mid \text{A wear}) \cdot P(\text{A wear}) + \\ &P(\text{a girl wear} \mid \text{S wear}) \cdot P(\text{S wear}) + \\ &P(\text{a girl wear} \mid \text{G wear}) \cdot P(\text{G wear}) + \\ &P(\text{a girl wear} \mid \text{Z wear}) \cdot P(\text{Z wear}) \end{aligned}}$$

$$= \frac{\frac{9}{14}}{1\cdot\frac{9}{14}+1\cdot 1+0\cdot 0+0\cdot\frac{4}{7}} = \frac{\frac{9}{14}}{\frac{23}{14}} = \frac{9}{23} \fallingdotseq \underline{0.391}$$

So the chance that it's Alexa is 0.391.

(c) $E[X_A] = \sum_{x\in X_A} x \cdot P[x]$

$= 2\cdot\frac{9}{14} + 1\cdot 0 + 0\cdot\frac{5}{14}$

$= \frac{9}{7} \fallingdotseq \underline{1.286 \cdots \text{Alexa}}$

$E[X_S] = \sum_{x\in X_S} x\cdot P[x]$

$= 1\cdot 1 = \underline{1 \cdots \text{Siri}}$

$E[X_G] = \sum_{x\in X_G} x\cdot P[x]$

$= 0\cdot 1 + 1\cdot 0 + 2\cdot 0 = \underline{0 \cdots \text{Googs}}$

$E[X_Z] = \sum_{x\in X_Z} x\cdot P[x]$

$= 2\cdot\frac{4}{7} + 0\cdot\frac{3}{7} = \frac{8}{7} \fallingdotseq \underline{1.143 \cdots \text{Zuckie}}$

(d) $V[x] = E[x^2] - E^2[x]$

$= E[(x - E[x])^2]$

$E[X_A^2] = \sum_{x\in X_A} x^2 P[x]$

$= 2^2\cdot\frac{9}{14} = \frac{18}{7}$

$E[X_S^2] = \sum_{x\in X_S} x^2 P[x]$

$= 1^2\cdot 1 = 1$

$E[X_G^2] = 0$

$E[X_Z^2] = \frac{16}{7}$

$V[X_A] = E[X_A^2] - (E[X_A])^2$

$= \frac{18}{7} - \left(\frac{9}{7}\right)^2 = \frac{126-81}{49} = \frac{45}{49} \fallingdotseq 0.918 \cdots \text{Alexa}$

$$V[X_S] = E[X_S^2] - (E[X_S])^2$$
$$= 1 - 1^2 = \underline{0 \cdots \text{Siri}}$$
$$V[X_G] = E[X_G^2] - (E[X_G])^2$$
$$= 0 - 0^2 = \underline{0 \cdots \text{Googs}}$$
$$V[X_Z] = E[X_Z^2] - (E[X_Z])^2$$
$$= \frac{16}{7} - \left(\frac{8}{7}\right)^2 = \frac{112-64}{49} = \frac{48}{49}$$
$$= 0.97959\ldots \approx \underline{0.980 \cdots \text{Zuckie}}$$

5. (a)

fur\tail	furry	rope-like
blue	0.1	0
gray	0.1	0
brown	0	0.8

$$P(\text{blue}, \text{furry}) = \frac{10 \cdot 1/2}{10+40} = \frac{5}{50} = \frac{1}{10} = 0.1$$
$$P(\text{blue}, \text{rope-like}) = \frac{0}{50} = 0$$
$$P(\text{gray}, \text{furry}) = \frac{10 \cdot 1/2}{50} = \frac{1}{10} = 0.1$$
$$P(\text{gray}, \text{rope-like}) = \frac{0}{50} = 0$$
$$P(\text{brown}, \text{furry}) = \frac{0}{50} = 0$$
$$P(\text{brown}, \text{rope-like}) = \frac{40}{50} = \frac{4}{5} = 0.8$$

(b) $P(\text{furry} = \text{blue}) = \frac{5}{50} = \frac{1}{10}$

$P(\text{tail} = \text{bushy}) = \frac{10}{50} = \frac{1}{5}$

$P(\text{furry} = \text{blue},\ \text{tail} = \text{bushy}) = \frac{1}{10}$

$\neq P(\text{furry} = \text{blue}) \cdot P(\text{tail} = \text{bushy})$

so $P(A, B) \neq P(A) \cdot P(B)$

they are not independent, they are dependent.

(C)

fur \ tail	furry	rope-like
blue	0.5	0
gray	0.5	0
brown	0	0

d) know that the hdding is Tom's kid, so

$P(\text{furry} = \text{brown}) = 0$, $P(\text{tail} = \text{rope-like}) = 0$.

$P(\text{tail} = \text{bushy}) = 1$. $P(\text{furry} = \text{brown}) = 0$

$P(\text{furry} = \text{blue}) = \frac{1}{2}$, $P(\text{furry} = \text{gray}) = \frac{1}{2}$

so $P(\text{furry} = \text{blue},\ \text{tail} = \text{bushy}) = P(\text{blue}) \cdot P(\text{bushy})$

$= \frac{1}{2}$

$P(\text{furry} = \text{gray},\ \text{tail} = \text{bushy}) = P(\text{gray}) \cdot P(\text{bushy})$

$= \frac{1}{2}$

so for any condition $P(A, B) = P(A) \cdot P(B)$

so they are Independent.

6. (a) $(1-e^{-\lambda x})' = 0 - e^{-\lambda x}\cdot(-\lambda) = \lambda e^{-\lambda x}$

so $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x>0. \\ 0 & . \ x \le 0. \end{cases}$

$$\int_0^x \lambda e^{-\lambda t} dt = \int_0^{\lambda x} e^{-t} dt$$
$$= -e^{-t}\Big|_0^{\lambda x}$$
$$= -e^{-\lambda x} - (-e^0)$$
$$= -e^{-\lambda x} + 1$$

b) $E[X] = \int_{-\infty}^{+\infty} x f(x)\, dx$

$$= \int_{-\infty}^{0} x f(x) dx + \int_0^{+\infty} x f(x) dx$$
$$= 0 + \int_0^{+\infty} x \lambda e^{-\lambda x} dx$$
$$= \frac{1}{\lambda}\int_0^{+\infty} \lambda x e^{-\lambda x} d\lambda x \quad (\text{let } \lambda x = t)$$
$$= \frac{1}{\lambda}\underline{\int_0^{+\infty} t e^{-t} dt}$$

$$\int_0^{+\infty} te^{-t} dt = \int_0^{+\infty} -t\, de^{-t}$$
$$= -\int_0^{+\infty} t\, de^{-t}$$

by $A\,dB = A\cdot B - B\,dA$

so $= -\left[te^{-t}\Big|_0^{+\infty} - \int_0^{+\infty} e^{-t} dt\right]$

$$= -\left[\frac{t}{e^t}\Big|_0^{+\infty} + e^{-t}\Big|_0^{+\infty}\right]$$
$$= -\left[\frac{t+1}{e^t}\Big|_0^{+\infty}\right]$$

By L'Hospital's rule:

$$\lim_{t\to+\infty}\frac{t+1}{e^t} = \lim_{t\to+\infty}\frac{(t+1)'}{(e^t)'} = \lim_{t\to+\infty}\frac{1}{e^t} = 0$$

so $-\left[\frac{t+1}{e^t}\Big|_0^{+\infty}\right] = -\left[0-\frac{1}{1}\right] = -(-1) = 1$

so $E(X) = \frac{1}{\lambda}\cdot 1 = \frac{1}{\lambda}$ $\quad V(X) = E(X^2) - (E(X))^2$

so $E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx$

$= 0 + \int_0^{+\infty} x^2 f(x) dx$

$= \int_0^{+\infty} x^2 f(x)\, dx$

$= \frac{1}{\lambda^2}\int_0^{+\infty} \lambda^2 x^2 e^{-\lambda x} d\lambda x$ (let $t = \lambda x$)

$= \frac{1}{\lambda^2} \underline{\int_0^{+\infty} t^2 e^{-t} dt}$

$\int_0^{+\infty} t^2 e^{-t} dt$

$= -\int_0^{+\infty} t^2 de^{-t}$

$= -[t^2 e^{-t} |_0^{+\infty} - \int_0^{+\infty} e^{-t} dt^2]$

$= -[\frac{t^2}{e^t}|_0^{+\infty} - 2\int_0^{+\infty} te^{-t} dt]$

By L'Hospital's rule:

$\int_0^{+\infty} t^2 e^{-t} dt = -[0 - 2\cdot 1]$

$= 2$

so $E(X^2) = \frac{1}{\lambda^2}\cdot 2 = \frac{2}{\lambda^2}$

so $V(X) = E(X^2) - (E(X))^2$

$\underline{= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}}$

so the mean is $1/\lambda$, the variance is $1/\lambda^2$

(C) Log likelihood

$\log p(D|\lambda)$. D means the things happened.

for the m observations:

$p(x_1|\lambda) = \lambda e^{-\lambda x_1}$

$p(x_2|\lambda) = \lambda e^{-\lambda x_2}$

... ...

$p(x_m|\lambda) = \lambda e^{-\lambda x_m}$

$$p(D|\lambda) = \prod_{i=1}^{m} \lambda e^{-\lambda x_i} \quad \cdots (\log A \cdot B = \log A + \log B$$

trans $\prod \to \sum$)

$$\ln p(D|\lambda) = \sum_{i=1}^{m} \ln \lambda + (-\lambda x_i)$$

the Maximum likelihood mean:

$$0 = \frac{\partial \ln p(D|\lambda)}{\partial \lambda}$$

$$= \sum_{i=1}^{m} \frac{1}{\lambda} - x_i$$

means $\frac{m}{\lambda_{MLE}} = \sum_{i=1}^{m} x_i$

means $\hat{\lambda}_{MLE} = m / \sum_{i=1}^{m} x_i$

$$\hat{\lambda} = \frac{m}{\sum_{i=1}^{m} x_i} = \frac{m}{x_1 + x_2 + \cdots + x_m}$$

(d) (i) yes.

$$\text{mean} = \frac{x_1 + x_2 + \cdots + x_m}{m}$$

$$= \sum_{i=1}^{m} x_i \, / m = \frac{1}{\hat{\lambda}}$$

$$E_D\left[\frac{1}{\hat{\lambda}}\right] = E_D\left[\sum_{i=1}^{m} x_i \, / m\right]$$

$$= \frac{1}{m} \sum_{i=1}^{m} E_D[x_i]$$

$$= \frac{1}{m} \cdot \sum_{i=m}^{m} \cdot \frac{1}{\lambda}$$

$$= \frac{1}{m} \cdot \left(m \cdot \frac{1}{\lambda}\right) = \frac{1}{\lambda}$$

so $1/\hat{\lambda}$ is unbiased estimate.

(ii) $E_D\left[\frac{1}{\hat{\lambda}^2}\right] = E_D\left[\left(\sum_{i=1}^{m} x_i \, / m\right)^2\right]$

$$= \frac{1}{m^2} E_D\left[\left(\sum_{i=1}^{m} x_i\right)^2\right]$$

$$= \frac{1}{m^2} E_D\left[(x_1 + x_2 + \cdots + x_m)^2\right]$$

$$= \frac{1}{m^2} E\left[x_1^2 + x_2^2 + \cdots + x_m^2 + 2x_1x_2 + 2x_1x_3 + \cdots + 2x_{m-1}x_m\right]$$

$$= \frac{1}{m^2} E\left[\sum_{i=1}^{m} x_i^2 + \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} 2 x_i x_j \right]$$

$$= \frac{1}{m^2} \left[\sum_{i=1}^{m} E(x_i^2) + 2 \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} E(x_i x_j) \right]$$

$$= \frac{1}{m^2} \left[m \cdot \frac{2}{\lambda^2} + 2 \cdot \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \frac{1}{\lambda^2} \right]$$

$$= \frac{1}{m^2} \left[m \cdot \frac{2}{\lambda^2} + 2 \cdot \left[(m-1+m-2+\cdots+1) \cdot \frac{1}{\lambda^2} \right] \right.$$

$$= \frac{1}{m^2} \left[2m \cdot \frac{1}{\lambda^2} + 2 \cdot \frac{(m-1)m}{2} \cdot \frac{1}{\lambda^2} \right]$$

$$= \frac{1}{m^2} \left[\frac{1}{\lambda^2} \cdot 2m + (m^2 - m) \frac{1}{\lambda^2} \right)$$

$$= \frac{1}{m^2} \left[\frac{2m-m}{\lambda^2} + \frac{m^2}{\lambda^2} \right)$$

$$= \frac{1}{m^2} \left[\frac{1}{\lambda^2} (m^2 + m) \right]$$

$$= \frac{1}{\lambda^2} \left(1 + \frac{1}{m} \right) \neq \frac{1}{\lambda^2}$$

<u>so $1/\hat{\lambda}^2$ is not an unbiased estimate.</u>

(e) by Hoeffding's inequality

$$\bar{X} = \frac{1}{n} (x_1 + x_2 + \cdots + x_n)$$

$$P(\bar{X} - E[\bar{X}] \geq t) \leq e^{-2nt^2} \quad \text{where } t > 0$$

also, by theorem 2 of Hoeffding:

$$P(\bar{x} - E[\bar{x}] \geq t) \leq \exp\left(-\frac{2n^2 t^2}{\sum_{i=1}^{n} (b_i - a_i)^2} \right.$$

$$P(|\bar{x} - E[\bar{x}]| \geq t) \leq 2\exp\left(-\frac{2n^2 t^2}{\sum_{i=1}^{n} (b_i - a_i)^2} \right)$$

$$\text{so } P(|\bar{x} - E[x]| \leq t) \geq 1 - 2\exp\left(-\frac{2n^2 t^2}{\sum_{i=1}^{n} (b_i - a_i)^2} \right)$$

now, $b_i = c$, $a_i = 0$. so we get

$$P(-t \leq \bar{x} - E[\bar{x}] \leq t) \geq 1 - 2\exp\left(-\frac{2n^2 t^2}{c^2} \right)$$

$$P(-\bar{x} - t \leq -E[\bar{x}] \leq -\bar{x} + t) \geq 1 - \delta$$

$$P(\bar{x} - t \leq E[\bar{x}] \leq \bar{x} + t) \geq 1 - \delta$$

and we have $P(\hat{\lambda}_{min} \le E[\bar{x}] \le \bar{x}+t) \ge 1-\delta$

so $\hat{\lambda}_{min} = \bar{x} - t$

$$= \frac{\sum_{i=1}^{n} x_i}{n} - \sqrt{-\frac{c^2}{2m}\ln\frac{\delta}{2}}$$

$$\hat{\lambda}_{max} = \bar{x} + t$$

$$= \frac{\sum_{i=1}^{n} x_i}{n} + \sqrt{-\frac{c^2}{2m}\ln\frac{\delta}{2}}$$

Challenge:

1. (a) $x^T y = \sum_{i=1}^{n} x_i y_i$

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

$$\|y\|_\infty = \max_i |y_i|$$

set $\|y\|_\infty = \max_i |y_i| = d \ge 0$

then for any i, $|y_i| \le d$.

$$x^T y = \sum_{i=1}^{n} x_i y_i$$

$$\le \sum_{i=1}^{n} |x_i y_i| = \sum_{i=1}^{n} |x_i||y_i|$$

$$\le \sum_{i=1}^{n} |x_i| \cdot d$$

$$= \|x\|_1 \|y\|_\infty$$

so $x^T y \le \|x\|_1 \|y\|_\infty$

(b) if $x = [0, 0, 0, \dots, 0]^T$

then $x^T y = 0$, $\|x\|_1 \|y\|_\infty = (\sum_{i=1}^{n} |0|) \cdot \|y\|_\infty$

$$= 0 \cdot \infty$$

for any $y \in R^n$, $x^T y = \|x\|_1 \|y\|_\infty$

set $x[i] \ne 0$, so the new set $L = \{l_1, l_2 \cdots l_m\}$.

which means that $\begin{cases} x[l_i] \neq 0, & i \in L \\ x[i] = 0. & i \notin L \end{cases}$

$$x^T y = \sum_{i=1}^{n} x_i y_i = \sum_{i\in L} x_i y_i + \sum_{i \notin L} x_i y_i$$

$$= \sum_{i \in L} x_i y_i + \sum_{i\notin L} 0 \cdot y_i$$

$$\|x\|_1 = \sum_{i}^{n} |x_i| = \sum_{i\in L} |x_i| + \sum_{i\notin L} |x_i|$$

$$= \sum_{i\in L} |x_i| + \sum_{i \notin L} 0 = \sum_{i \in L} |x_i|$$

$$x^T y = \|x\|_1 \|y\|_\infty \Rightarrow \sum_{i\in L} x_i y_i = \sum_{i \in L} |x_i| \cdot \max_i |y_i|$$

$$\Rightarrow \max_i |y_i| = \frac{\sum_{i\in L} x_i y_i}{\sum_{i\in L} |x_i|}$$

2. (a) $x^T y = \sum_i x_i y_i$

we can set $\max_i y_i = d$

then for any i, $y_i \leq d$.

then $(\sum_i x_i) \max_i y_i = d(\sum_i x_i)$

$(\sum_i x_i) \max_i y_i - x^T y$

$= \sum_i x_i \cdot d - \sum_i x_i y_i$

$= \sum_i (d x_i) - \sum_i x_i y_i$

$= \sum_i (d x_i - x_i y_i) = \sum_i (d - y_i) x_i$

for any i $\begin{cases} d - y_i \geq 0 \; (y_i \leq d) \\ x_i \geq 0 \end{cases}$

so $(d - y_i) x_i \geq 0$

so $\sum_i (d - y_i) x_i \geq 0$

so $(\sum_i x_i) \max_i y_i - x^T y \geq 0$