Challenge

Q3) $E[x] = \sum_{i=1}^{n} x_i p_i$

$E[f(x)] = \sum_{i=1}^{n} p_i f(x_i)$

$$f(E[x]) = f(x_1 p_1 + x_2 p_2 \dots x_n p_n)$$

$$f(E(x)) = f\left(p_1 x_1 + (1-p_1)\left[\frac{p_2}{1-p_1} x_2 + \frac{p_3}{1-p_1} x_3 \dots \frac{p_n}{1-p_n} x_n\right]\right)$$

$$< p_1 f(x_1) + (1-p_1)\, f\left[\frac{p_2}{1-p_1} x_2 + \frac{p_3}{1-p_1} x_3 \dots \frac{p_n}{1-p_n} x_n\right]$$

(Assuming $\theta = p_1$)

$$f(E[x]) < p_1 f(x_1) + (1-p_1)\, f\left[\frac{p_2}{1-p_1} x_2 + \frac{p_3}{1-p_1} x_3 \dots \frac{p_n}{1-p_n} x_n\right] \quad \text{— (1)}$$

Let us consider the term,

$$= f\left[\frac{p_2}{1-p_1} x_2 + \frac{p_3}{1-p_1} x_3 \dots \frac{p_n}{1-p_n} x_n\right]$$

$$= f\left(\frac{p_2 x_2}{1-p_1} + \frac{1-p_1-p_2}{1-p_1}\left[\frac{p_3 x_3}{1-p_1-p_2} + \frac{p_4 x_4}{1-p_1-p_2} \dots + \frac{p_n x_n}{1-p_1-p_2}\right]\right)$$

$$< \frac{P_2}{1-P_1} f(x_2) + \frac{1-P_1-P_2}{1-P_1} f\left[\frac{P_3x_3}{1-P_1-P_2} + \frac{P_4x_4}{1-P_1-P_2} \cdots + \frac{P_nx_n}{1-P_1-P_2}\right]$$

$$\theta = \frac{P_2}{1-P_1} \qquad 1-\theta = 1-\frac{P_2}{1-P_1} = \frac{1-P_1-P_2}{1-P_1}$$

$$\Rightarrow (1-P_1) f\left(\frac{P_2x_2}{1-P_1} + \frac{P_3x_3}{1-P_1} \cdots \frac{P_nx_n}{1-P_1}\right)$$

$$< P_2 f(x_2) + (1-P_1-P_2) f\left(\frac{P_3x_3}{1-P_1-P_2} \cdots + \frac{P_nx_n}{1-P_1-P_2}\right) \quad — (2)$$

From (1) & (2), we get,

$$f(E[x]) < P_1 f(x_1) + P_2 f(x_2) + (1-P_1-P_2) f\left(\frac{P_3x_3}{1-P_1-P_2} \cdots + \frac{P_nx_n}{1-P_1-P_2}\right)$$

If we generalize the above term, we get

$$f(E[x]) < P_1 f(x_1) + P_2 f(x_2) + P_3 f(x_3) \cdots + P_{n-1} f(x_{n-1}) + (1-P_1-P_2\cdots-P_{n-1}) f\left(\frac{P_nx_n}{1-P_1-P_2\cdots-P_{n-1}}\right)$$

$\Rightarrow p_1 + p_2 + p_3 \ldots \quad p_n = 1$

$\Rightarrow 1 - p_1 - p_2 - p_3 \ldots - p_{n-1} = p_n$

$f(E[x]) \leq p_1 f(x_1) + p_2 f(x_2) \ldots$

$+ p_{n-1} f(x_{n-1}) + p_n f(x_n)$

$$\boxed{\therefore \; f(E[x]) \leq E[f(x)]}$$

Q4) $P_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{else} \end{cases}$

(a) Likelihood (L),

$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}$

$= \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$

$\log(L(\lambda)) = n \log \lambda - \lambda \sum_{i=1}^{n} x_i$

For maximum, set derivative to '0'.

$$\frac{dL(\lambda)}{d\lambda} = \frac{n}{\lambda} - \Sigma x_i = 0$$

$$\boxed{\lambda_{MLE} = \frac{n}{\Sigma x_i} \qquad \lambda_{MLE} = \frac{1}{\bar{\theta}} \quad \left(\bar{\theta} = \frac{\sum_{i=1}^{n} x_i}{n}\right)}$$

b) $$E[\hat{\lambda}] = E\left[\frac{n}{\sum_{i=1}^{n} x_i}\right]$$

$$= n \times E\left[\frac{1}{x_i}\right]$$

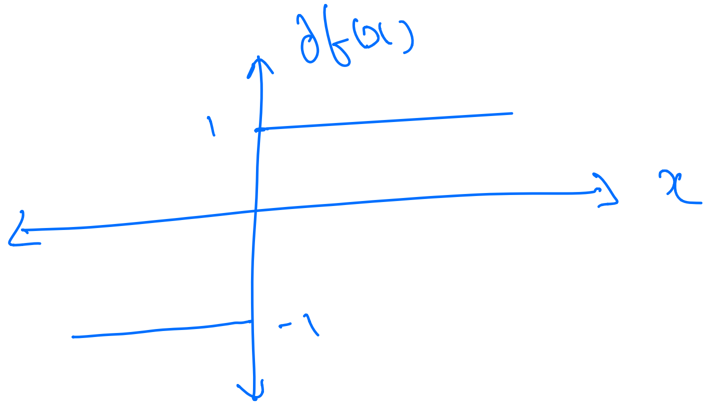
$$E[x^{-1}] = \int_0^\infty \frac{1}{\lambda} \lambda e^{-\lambda x} dx = \frac{\lambda}{n-1}$$

$$\boxed{E(\hat{\lambda}) = \frac{n}{n-1} \lambda \neq \lambda}$$ $\rightarrow$ It is a biased estimator

Q2)

(a) If f is convex and differentiable, then gradient at x is a sub gradient.

But a sub-gradient can exist even when f is non-differentiable.



A function f is called sub-differentiable at x if there exists atleast one sub-gradient at x.

Consider, $f(x) = |x|$

For, $x < 0 \rightarrow$ sub-gradiet $\partial f(x) = -1$

For, $x > 0 \longrightarrow$ sub-gradiet $\partial f(x) = 1$

At $\boxed{x = 0}$

one sub-gradient is defined by the equality, $\boxed{|z| \geq gz}$ $\forall z$ which is satisfied iff $\boxed{g \in [-1, 1]}$.

$\therefore \partial f(0) = [-1, 1]$

$$\therefore \partial f(x) = \begin{cases} \{1\} & \text{if } x > 0 \\ \{-1\} & \text{if } x < 0 \\ [-1, 1] & x = 0 \end{cases}$$

b) A point x^* is a minimizer of a convex function iff f is sub-differentiable at x^* and $0 \in \partial f(x^*)$,

i.e. $g = 0$ is a sub-gradient of f at x^*.

$[\text{as } f(x) \geq f(x^*)]$

$\Rightarrow 0 \in \partial f(x^*)$ reduces to $\nabla f(x^*) = 0$ if f is differentiable at x^*.