Q1) The neg. log likelihood is

$$L(\{x_i\}_{i=1}^{n} \mid \mu) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2/2\sigma^2$$

$$\longrightarrow \hat{\mu} = \sum_i x_i/n$$

a) $v = a\mu+b \rightarrow \mu = (v-b)/a$

$$\therefore \mathcal{L}(\{x_i\} \mid v) = \frac{1}{n}\sum_i \left(x_i - \frac{v-b}{a}\right)^2/2\sigma^2$$

$$\therefore \partial L/\partial v = \frac{1}{n}\sum_{i=1}^{n}\frac{\cancel{2}\left(x_i-\frac{v-b}{a}\right)}{\cancel{2}\sigma^2}\cancel{\left(\frac{-1}{a}\right)} = 0$$

$$\therefore \hat{v} = \frac{1}{n}\left(\sum_{i=1}^{n} a x_i\right) + b = a\hat{\mu}+b = g(\hat{\mu})$$

$$E(\hat{v}) = aE(\hat{\mu})+b = a\mu+b = g(\mu)$$

So this is an unbiased estimator

b) $v = \mu^3 \longrightarrow \mu = v^{1/3}$

$$\mathcal{L}(\{x_i\} \mid v) = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - v^{1/3}\right)^2/2\sigma^2$$

$$\partial L/\partial v = \cancel{\frac{1}{n}}\sum_i \frac{\cancel{2}(x_i - v^{1/3})}{\cancel{2}\sigma^2}\,\cancel{\frac{1}{3}}\,v^{-2/3} = 0$$

As $v \neq 0$, $\hat{v} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^3 = \hat{\mu}^3 = g(\hat{\mu})$

(2)

$E(\hat{\mu}^3) = ?$ $E(\hat{\mu}) = \mu$, $E[\hat{\mu}^2] = Var(\hat{\mu}) - (E(\hat{\mu}))^2$

$E[(\hat{\mu} - \mu)^3] = 0$ as $\hat{\mu}$ is Gaussian distributed

$$E[\hat{\mu}^3 - \mu^3 + 3\hat{\mu}\mu^2 - 3\hat{\mu}^2\mu] = 0$$

$$E[\hat{\mu}^3] - \mu^3 + 3\mu^2\mu - 3\mu E[\hat{\mu}^2] = 0$$

$$E[\hat{\mu}^3] + 2\mu^3 - 3\mu[Var(\hat{\mu}) + \mu^2] = 0$$

$$E[\hat{\mu}^3] + 2\mu^3 - 3\mu\left[\frac{\sigma^2}{n}\right] - 3\mu^3 = 0$$

$$E[\hat{\mu}^3] = \mu^3 + 3\mu\frac{\sigma^2}{n} \neq \mu^3$$

So this is not an unbiased estimator

Q2) $$P(A|x) = p(x|A)\,P(A) / p(x)$$

$$= \frac{\dfrac{e^{-(x-1)^2/2}}{\sqrt{2\pi}} \, 1/3}{p(x)}$$

$$P(B|x) = \frac{p(x|B)\,P(B)}{p(x)} = \frac{\dfrac{e^{-(x-2)^2/4} \times \frac{2}{3}}{\sqrt{2\pi}\;2}}{p(x)}$$

For values of x such that $P(A|x) = P(B|x)$, ③
its is impossible to do the classification using conditionals on x alone.

ie for which

$$\frac{e^{-\frac{(x-1)^2}{2}}}{\cancel{3}} = \frac{e^{-\frac{(x-2)^2}{4}}}{\cancel{3}}$$

$$\rightarrow \frac{(x-1)^2}{\cancel{2}} = \frac{(x-2)^2}{\cancel{4}\,2}$$

$$\rightarrow 2[x^2 - 2x + 1] = x^2 - 4x + 4$$

$$\rightarrow 2x^2 - \cancel{4}x + 2 = x^2 - \cancel{4}x + 4 \rightarrow x = \pm\sqrt{2}$$

Q3) See class lecture slides.

Q4) $Y = 1/x$

$$P(Y \le y) = P(X \geq 1/y) = 1 - P(X < 1/y)$$

$$= 1 - F_X(1/y)$$

$$\downarrow$$

$$F_Y(y)$$

$$\therefore f_Y(y) = -f_X(1/y)\left(\frac{-1}{y^2}\right) = \frac{f_X(1/y)}{y^2}$$

$$f_X(x) = \begin{cases} \frac{1}{b-a} & 0 \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

$$\therefore f_Y(y) = \begin{cases} \frac{1}{y^2(b-a)} & 1/b \le y \le 1/a \\ 0 & \text{otherwise} \end{cases}$$

$$F_Y(y) = 1 - F_X(1/y)$$
$$= 1 - \frac{1/y - a}{b-a} = \frac{b - 1/y}{b-a}$$

$$E(Y) = \int_{1/b}^{1/a} \frac{1}{y^2(b-a)} y \, dy = \frac{(\ln y)_{1/b}^{1/a}}{b-a}$$
$$= \frac{\log(1/a) - \log(1/b)}{b-a} = \frac{\ln b - \ln a}{b-a}$$

$F_Y(y) = 1/2 \rightarrow y$ is median

$$\frac{b - 1/y}{b-a} = \frac{1}{2} \longrightarrow y = \frac{2}{a+b}$$

$$Var(Y) = E(Y^2) - (E(Y))^2$$
$$E(Y^2) = \int_{1/b}^{1/a} \frac{1}{y^2} \frac{y^2}{b-a} dy = \frac{\frac{1}{a} - \frac{1}{b}}{b-a} = \frac{1}{ab}$$

(5)

$$Var(Y) = \frac{1}{ab} - \left(\frac{\ln b - \ln a}{b-a}\right)^2$$

——— X ———

Q5

$$P(X \geq \lambda + a) = P(e^{tX} \geq e^{t(\lambda+a)})$$

$$= P(e^{t(X-\lambda-a)} \geq 1) \leq E[e^{t(X-\lambda-x)}]$$

by Markov's inequality

$$= e^{\lambda(e^t-1)} e^{-\cancel{t}(\lambda+x)} = RHS$$

$$e^{\lambda(e^t-1)-t(\lambda+x)}$$

$$\frac{\partial RHS}{\partial t} = 0 \longrightarrow e^{\lambda(e^t-1)-t(\lambda+x)} (\lambda e^t - \lambda - x) = 0$$

$$\longrightarrow e^t = 1 + x/\lambda \longrightarrow t = \ln\left(1+\frac{x}{\lambda}\right)$$

$\therefore$ RHS is minimized when $t = \ln(1 + x/\lambda)$

$$\therefore \min RHS = e^{\lambda\left[1+\frac{x}{\lambda}-1\right] - \ln\left(1+\frac{x}{\lambda}\right)(\lambda+x)}$$

$$= e^{x - \ln(1+x/\lambda)(\lambda+x)}$$

$$-\frac{x^2}{2\lambda} h\left(\frac{x}{\lambda}\right) = \frac{-\cancel{x^2}}{\cancel{2\lambda}} \times \frac{\cancel{2}\left((1+x/\lambda)\ln(1+x/\lambda) - x/\lambda\right)}{\cancel{x^2/\lambda^2}}$$

$$= \frac{-1}{\cancel{\lambda}}\left[(\lambda+x)\ln(1+x/\lambda) - x\right] = -RHS$$

$\therefore$ proved

6

For the second inequality, we have

$$P(X \le \lambda - x) = P\left(e^{tX} \le e^{t(\lambda-x)}\right)$$

$$= P\left(e^{t(\lambda - x - X)} \ge 1\right)$$

$$\le E\left[e^{t(\lambda-x-X)}\right]$$ by Markov's inequality

$$= e^{t(\lambda-x)} E\left(e^{-tX}\right)$$

$$= e^{t(\lambda-x)} e^{\lambda(e^{-t}-1)}$$

$$= e^{t(\lambda-x)+\lambda(e^{-t}-1)}$$

RHS is minimized for t given as

$$e^{t(\lambda-x)+\lambda(e^{-t}-1)}\left(\lambda - x - \lambda e^{-t}\right) = 0$$

$$\longrightarrow \lambda e^{-t} = \lambda - x \longrightarrow e^{-t} = 1 - x/\lambda$$

$$\longrightarrow t = -\log(1 - x/\lambda)$$

Plugging in, the optimal RHS

$$= e^{-\log(1-x/\lambda)(\lambda - x) + \lambda\left(\cancel{1} - \frac{x}{\lambda} - \cancel{1}\right)}$$

$$\frac{-x^2}{2\lambda} h\left(-\frac{x}{\lambda}\right) = \frac{-\cancel{x^2}}{\cancel{2}\cancel{\lambda}} \cancel{2} \frac{\left(1-\frac{x}{\lambda}\right)\log\left(1-\frac{x}{\lambda}\right) + \frac{x}{\lambda}}{\cancel{x^2}/\cancel{\lambda^2}}$$

$$= -(\lambda - x)\log\left(1 - \frac{x}{\lambda}\right) - x = \text{RHS} \therefore \text{proved}$$

NegJLL

$$= \sum_{i \in S_1} \left[\frac{(x_i - \mu_1)^2}{2\sigma^2} + \log \sigma \right]$$

$$+ \sum_{i \in S_2} \left[\frac{(x_i - \mu_2)^2}{2\sigma^2} + \log \sigma \right] + \cdots + \sum_{i \in S_k} \left[\frac{(x_i - \mu_k)^2}{2\sigma^2} + \log \sigma \right]$$

$\gamma = \sigma^2$

$$\frac{\partial NJLL}{\partial \sigma} = \sum_{i \in S_1} \frac{(x_i - \mu_1)^2 (-\cancel{2})}{\cancel{2}\, \sigma^3} + \cdots + \sum_{i \in S_k} \frac{(x_i - \mu_k)^2 (-\cancel{2})}{\cancel{2}\sigma^3}$$

$$+ \frac{(n_1 + n_2 + \cdots + n_k)}{\sigma} = 0$$

$$\sum_{i \in S_1} \frac{(x_i - \mu_1)^2}{\sigma^2} + \cdots + \sum_{i \in S_k} \frac{(x_i - \mu_k)^2}{\sigma^2} = \frac{n}{1}$$

$$\therefore \sigma^2 = \frac{1}{n} \left[\sum_{i \in S_1} (x_i - \mu_1)^2 + \cdots + \sum_{i \in S_k} (x_i - \mu_k)^2 \right]$$

As $\mu_1, \cdots \mu_k$ are unknown we have

$$\hat{\sigma}^2 = \frac{1}{n} \left[\sum_{i \in S_1} (x_i - \bar{x}_1)^2 + \cdots + \sum_{i \in S_k} (x_i - \bar{x}_k)^2 \right]$$

$$E\left[\sum_i (x_i - \bar{x}_1)^2\right] = E\left[\sum_i x_i^2 + \bar{x}_1^2 - 2\bar{x}_1 x_i\right]$$

$$= n_1(\mu_1^2 + \sigma^2) + n\mu_1^2 + \frac{\sigma^2 n}{n} - 2\cancel{n}\frac{\mu_1}{\cancel{n}}\mu_1$$

$$E\left[\sum_{i\in S_1}(x_i-\overline{x}_1)^2\right]$$

$$= E\left[\sum_{i\in S_1}\left(x_i^2+\overline{x}_1^{\,2}-2x_i\overline{x}_1\right)\right]$$

$$= n_1(\mu_1^2+\sigma^2)+n_1E(\overline{x}_1^{\,2})-2E\Big(\overline{x}_1\sum_{i\in S_1}x_i\Big)$$

$$= n_1(\mu_1^2+\sigma^2)+n_1\left[\mu_1^2+\frac{\sigma^2}{n_1}\right]-2n_1E(\overline{x}_1^{\,2})$$

$$= n_1(\mu_1^2+\sigma^2)-n_1\left(\mu_1^2+\frac{\sigma^2}{n_1}\right)$$

$$= \cancel{n_1\left[\sigma^2-1\right]}\ (n_1-1)\,\sigma^2$$

$$\therefore E(\hat{\sigma}^2)=\frac{1}{n}\sigma^2\left[n_1+n_2+\cdots+n_k-k\right] \longrightarrow (1)$$

$\therefore$ $\hat{\sigma}^2$ is a biased estimator.

Now if $\mu_1, \mu_2, \ldots$ are known but μ_{k-1} and μ_k are not known, then

$\overline{x}_{k-1}$

$$\hat{\sigma}^2=\frac{1}{n}\left(\sum_{l=1}^{k-2}\sum_{i\in S_l}(x_i-\mu_l)^2+\sum_{i\in S_{k-1}}(x_i-\mu_{k-1})^2\right.$$

$$\left.+\sum_{i\in S_k}(x_i-\mu_k)^2\right]$$

$\overline{x}_k$

$$E(\hat{\sigma}^2) = \frac{1}{n}\sigma^2\left[n_1+n_2+\cdots+n_k-2\right]$$ (9)

This is still a biased estimate. → (2)

To correct (1) for bias, we multiply

$E(\hat{\sigma}^2)$ with $\dfrac{n}{n_1+n_2+\cdots+n_k-k}$.

To correct (2) for bias, we multiply

$E(\hat{\sigma}^2)$ with $\dfrac{n}{n_1+n_2+\cdots+n_k-2}$.

———— X ————

Q7 a) $E(F_n(x)) = \dfrac{1}{n}\sum_i \underline{E(1(X_i \le x))}$

$= \dfrac{1}{n} \times n\, F(x)$ as X_i are iid and $1(X_i \le x)$ is a Bernoulli r.v. with parameter $F(x)$

$= F(x)$

So unbiased estimator. bias $= 0$

$$Var(F_n(x)) = \frac{1}{n^2}\sum_i Var(1(X_i \le x))$$

$$= \frac{1}{n^2} \times n\, F(x)(1-F(x)) = \frac{F(x)(1-F(x))}{n}$$

$$MSE = \frac{F(x)(1-F(x))}{n}$$

b) by CI,

$$P_r\left(|F_n(x) - F(x)| \geq k\, var\right) \leq \frac{1}{k^2}$$

$$\therefore P\left(|F_n(x) - F(x)| \geq \varepsilon\right) \leq \frac{Var}{\varepsilon^2}$$

$$= \frac{F(x)(1-F(x))}{n\varepsilon^2}.$$

c) $F_n(x)$ is approx. Gaussian distributed with mean $F(x)$ and variance $F(x)(1-F(x))/n$ via CLT.

$$\therefore P\left(\frac{|F_n(x) - F(x)|}{\sqrt{F(x)(1-F(x))/n}} \geq \varepsilon\right) \leq \frac{e^{-\varepsilon^2/2}}{\varepsilon\sqrt{2\pi}}$$ by Gaussian tail bound

$$\therefore P\left(|F_n(x) - F(x)| \geq \varepsilon\right) \leq \frac{e^{-\varepsilon^2/2}\sqrt{F(x)(1-F(x))}}{\varepsilon\sqrt{2\pi}\,\sqrt{n}}$$

d) The bound in c) is tighter but relies on Gaussian approximation which holds only given many (infinite) samples

e) We have

$$P\left(\max_x |F_n(x) - F(x)| \geq \varepsilon\right) \leq 2e^{-2n\varepsilon^2} \quad \text{by DKW}$$

$$\therefore P\left(\forall x \ |F_n(x) - F(x)| \leq \varepsilon\right) \geq 1 - 2e^{-2n\varepsilon^2}$$

$$\cancel{\therefore P\left(\forall x \ |F_n(x) - F(x)|\right.}$$

Let $\alpha = 2e^{-2n\varepsilon^2}$ ie

$$-2n\varepsilon^2 = \log(\alpha/2)$$

$$\therefore \varepsilon = \sqrt{\frac{\log(2/\alpha)}{2n}}$$

$$\therefore L(x) = \max\left[\hat{F}_n(x) - \sqrt{\frac{\log(2/\alpha)}{2n}}, 0\right]$$

$$U(x) = \min\left[\hat{F}_n(x) + \sqrt{\frac{\log(2/\alpha)}{2n}}, 1\right]$$

as $0 \leq F(x) \leq 1$

Then $P(\forall x \in \mathbb{R}, \ L(x) \leq F(x) \leq U(x)) \geq 1 - \alpha$