Q1

Solution. To compute the MLE for the numerator $\sup_{\theta\in\Theta} f_\theta(x)$ where $x = (x_1, ..., x_n) \in \mathbb{R}^n$, the joint PDF is

$$\prod_{i=1}^n 1_{\{x_i \geqslant \theta\}} e^{-(x_i-\theta)} = 1_{\{x_{(1)} \geqslant \theta\}} \exp(-\sum_{i=1}^n X_i) \exp(n\theta).$$

It is clear that we want θ as large as possible, subject to the constraint $\theta \leqslant x_{(1)}$. Hence

$$\sup_{\theta\in\Theta} f_\theta(x) = \exp(-\sum_{i=1}^n x_i) \exp(nx_{(1)}).$$

For the denominator, the likelihood increases as θ increases, subject to both $\theta \leqslant \theta_0$ and $\theta \leqslant x_{(1)}$, i.e., the supremum is attained when $\theta \leqslant \min(\theta_0, x_{(1)})$. Hence

$$\frac{\sup_{\theta\in\Theta} f_\theta(x)}{\sup_{\theta\in\Theta_0} f_\theta(x)} = \frac{\exp(nx_{(1)})}{\exp(n\min(\theta_0, x_{(1)}))} = \exp[n(x_{(1)} - \min(\theta_0 - x_{(1)}))].$$

From this, since the region is defined by ratio $\geqslant k$, we have

$$x \in \text{rejection region} \iff \begin{cases} x_{(1)} \geqslant \dfrac{\log k}{n} + \theta_0 & \text{if } x_{(1)} \geqslant \theta_0 \\ 1 \geqslant \log k & \text{if } x_{(1)} < \theta_0. \end{cases}$$

Since k is assumed to be positive (or > 1), this reduces to

$$x \in \text{ rejection region } \iff x_{(1)} \geqslant \frac{\log k}{n} + \theta_0.$$

According to a theorem shown in class,

$$2\log\frac{\sup_{\theta\in\Theta} f_\theta(X)}{\sup_{\theta\in\Theta_0} f_\theta(X)}$$

converges in distribution to χ_1^2 as $n \to \infty$.

Q2

Solution. From a result (4.41) shown in class, the MLE for μ over all $\mu \in \mathbb{R}$ is the sample mean, and from another result (5.24), the rejection is given by

$$\left\{x \in \mathbb{R}^n : \left|\frac{1}{n}\sum_{i=1}^n x_i - \mu_0\right| \geqslant sqr\right\}$$

Problem 1

Suppose $X_1, ..., X_n$ is a random sample from a Gaussian random variable X with known variance $\mu_X \in \mathbb{R}$ and unknown variance $\sigma^2 > 0$. Suppose $Y_1, ..., Y_m$ is a random sample from Gaussian random variable with known mean $\mu_Y \in \mathbb{R}$ and unknown variance $\sigma^2 > 0$. Assume that X, Y are independent and that $m + n > 2$. Define

$$\overline{X} := \frac{1}{n}\sum_{i=1}^n X_i \qquad \overline{Y} := \frac{1}{m}\sum_{i=1}^m Y_i,$$

$$S_X^2 := \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2, \qquad S_y^2 := \frac{1}{m}\sum_{i=1}^{m}(Y_i - \overline{Y})^2,$$

and

$$S^2 := \frac{(n-1)S_X^2 + (m-1)S_Y^2}{m+n-2}.$$

Show that $\frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{S\sqrt{n^{-1} + m^{-1}}}$ has Student's t-distribution with $m+n-2$ degrees of freedom. Deduce that

$$\mathbb{P}(\overline{X} - \overline{Y} - tS\sqrt{n^{-1}+m^{-1}} < \mu_X - \mu_Y < \overline{X} - \overline{Y} + tS\sqrt{n^{-1}+m^{-1}}) = \frac{\Gamma((p+1)/2)}{\sqrt{p}\sqrt{\pi}\Gamma(p/2)}\int_{-t}^{t}\left(1+\frac{t^2}{p}\right)^{-(p+1)/2} \mathrm{d}t$$

where $p = m+n-2$.

Proof. Recall from definition that if Z is standard Gaussian, then

$$\sqrt{p}Z/\sqrt{\chi_p^2} \sim t_p. \tag{1}$$

Also recall that if $\tilde{S} := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ then

$$(n-1)\tilde{S}^2/\sigma^2 \sim \chi_{n-1}^2. \tag{2}$$

Finally, if χ_p^2 and χ_q^2 are independent, then

$$\chi_p^2 + \chi_q^2 \sim \chi_{p+q}^2. \tag{3}$$

By (1), we want to show that $S^2 \sim \chi_p^2$ where $p = m+n-2$. First notice that

$$S_X^2 = \frac{n-1}{n}\tilde{S}_X = \frac{n-1}{n}\frac{\sigma^2}{n-1}\chi_{n-1}^2 = \frac{\sigma^2}{n}\chi_{n-1}^2 \tag{4}$$

and similarly

$$S_Y^2 = \frac{\sigma^2}{m}\chi_{m-1}^2. \tag{5}$$

Applying (4) and (5) gives

$$(n-1)S_X^2 + (m-1)S_Y^2 = \left(\frac{n-1}{n}\right)\sigma^2\chi_{n-1}^2 + \left(\frac{m-1}{m}\right)\sigma^2\chi_{m-1}^2. \tag{6}$$

By independence of X and Y, we have independence of S_X^2 and Y_X^2. In the case that $m = n$,

$$S^2 = \frac{((n-1)/n)\sigma^2\chi_{2n-2}^2}{2n-2} \qquad \text{and} \qquad \frac{(\overline{X} - \overline{Y} - \mu_X + \mu_Y)\sqrt{n}}{\sigma\sqrt{2}} \sim \mathcal{N}(0,1).$$

Then,

$$\frac{\sqrt{2n-2}Z}{\sqrt{\chi_{2n-2}^2}} = \sqrt{2n-2}\cdot\frac{(\overline{X} - \overline{Y} - \mu_X + \mu_Y)\sqrt{n}}{\sigma\sqrt{2}}\cdot\left(\frac{\sigma^2(n-1)}{(2n-2)S^2 n}\right)^{1/2} = \frac{\overline{X} - \overline{Y} - \mu_X + \mu_Y}{S\sqrt{2/(n-1)}}?$$

□

Q5

Solution. The key observation is that if $X_1, \ldots, X_n$ are i.i.d. Gaussians, then

$$(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}, \qquad \text{where } S^2 := \frac{1}{n-1}(X_i - \overline{X})^2 \text{ and } \overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Hence, to construct a confidence interval for σ^2, we start with one for $(n-1)S^2/\sigma^2$. Let α, β be such that

$$\mathbb{P}(\chi^2_{n-1} \leqslant \alpha) = 0.025 \qquad \text{and} \qquad \mathbb{P}(\chi^2_{n-1} \geqslant \alpha) = 0.025.$$

Then $\mathbb{P}(\alpha \leqslant \chi^2_{n-1} \leqslant \beta) = \mathbb{P}(\alpha \leqslant (n-1)S^2/\sigma^2 \leqslant \beta) = 0.05$; that is,

$$\mathbb{P}\left(\frac{\alpha}{(n-1)S^2} \leqslant 1/\sigma^2 \leqslant \frac{\beta}{(n-1)S^2}\right) = \mathbb{P}\left(\frac{(n-1)S^2}{\beta} \leqslant \sigma^2 \leqslant \frac{(n-1)S^2}{\alpha}\right) = 0.05.$$

Therefore, $[(n-1)S^2/\beta, (n-1)S^2/\alpha]$ is our desired 95% confidence interval.
In our example, $n = 6$, so upon checking the table, $\alpha = 0.484$, $\beta = 0.831$, and $S^2 = 2$. Hence the answer is

$$[5 \cdot 2/0.831, 5 \cdot 2/0.484] \approx [12.034, 20.661].$$

Q6

Solution. Recall that if $X_1, \ldots, X_n$ are i.i.d. $\mathcal{N}(\mu_X, \sigma_X^2)$, if $Y_1, \ldots, Y_m$ are i.i.d. $\mathcal{N}(\mu_Y, \sigma_Y^2)$, and if X, Y are independent, then

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{m}\sum_{j=1}^{m} Y_j - (\mu_X - \mu_Y)$$

is normal with mean 0 and variance $\sigma_X^2/n + \sigma_Y^2/m$. Thus,

$$\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{m}\sum_{j=1}^{m} Y_j - (\mu_X - \mu_Y)}{\sqrt{\sigma_X^2/n + \sigma_Y^2/m}} \sim \mathcal{N}(0,1).$$

Thus a 95% confidence interval corresponds to the ratio lying in $[-2, 2]$; that is,

$$\mathbb{P}\left(\overline{X} - \overline{Y} - 2\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} \leqslant \mu_X - \mu_Y \leqslant \overline{X} - \overline{Y} + 2\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right) = 0.95.$$

In Q6, $\overline{X} = 2$, $\overline{Y} = 4$, $\sigma_X^2 = 2$, and $\sigma_Y^2 = 3$. Our interval is $[-2 - 2\sqrt{5/2}, -2 + 2\sqrt{5/2}] = [-2 - \sqrt{10}, -2 + \sqrt{10}]$.