1. a) $E[A] = E[2X+Y] = 2E[X] + E[Y] = \boxed{2\lambda + \frac{(a+b)}{2}}$

$E[B] = E[X-2Y] = E[X] - 2E[Y] = \lambda - (a+b) = \boxed{\lambda - a - b}$

b) $\text{var}(A) = E[(A-\mu_A)^2] = E[A^2] - 2\mu_A^2 + \mu_A^2 = E[A^2] - (E[A])^2$

$E[A^2] = E[(2X+Y)(2X+Y)] = E[4X^2 + 4XY + Y^2] = 4E[X^2] + 4\mu_X\mu_Y + E[Y^2]$

$E[X^2] = \lambda^2 + \lambda$

$E[Y^2] = \int_a^b y^2 \frac{1}{b-a}\, dy = \frac{y^3}{3}\left(\frac{1}{b-a}\right)\Big|_a^b = \frac{b^3}{3}\left(\frac{1}{b-a}\right) - \frac{a^3}{3}\left(\frac{1}{b-a}\right)$

$= \left(\frac{1}{b-a}\right)\left(\frac{b^3-a^3}{3}\right)$

$\text{var}(A) = E[A^2] - (E[A])^2$

$= \boxed{4(\lambda^2+\lambda) + 4\lambda\left(\frac{1}{b-a}\right) + \left(\frac{1}{b-a}\right)\left(\frac{b^3-a^3}{3}\right) - \left(2\lambda + \frac{a+b}{2}\right)^2}$

$\text{var}(A) = \text{Var}(2X+Y) = 4\,\text{Var}(X) + \text{Var}(Y) = \boxed{4\lambda + \frac{(b-a)^2}{12}}$ or

$\text{var}(B) = E[B^2] - (E[B])^2$

$E[B^2] = E[(X-2Y)^2] = E[X^2 - 4YX + 4Y^2] = E[X^2] - 4\mu_X\mu_Y + 4E[Y^2]$

$= \lambda^2 + \lambda - 4\lambda\left(\frac{a+b}{2}\right) + 4\left(\frac{1}{b-a}\right)\left(\frac{b^3-a^3}{3}\right)$

$\boxed{\text{var}(B) = \lambda^2 + \lambda - 4\lambda\left(\frac{a+b}{2}\right) + 4\left(\frac{1}{b-a}\right)\left(\frac{b^3-a^3}{3}\right) - (\lambda - a - b)^2}$

$\text{var}(B) = \text{Var}(X-2Y) = \text{Var}(X) + 4\text{Var}(Y) = \boxed{\lambda + \frac{(b-a)^2}{3}}$ or

c) Covariance, $E[(X-\mu_X)(Y-\mu_Y)]$ is a measure of how the combined average of X and Y differs from the piece-wise average. This measures the combined variance of X and Y. Correlation is a measure of how two random variable's distributions are related. A positive correlation indicates that the larger values of X correspond to the larger values of Y, while a negative correlation indicate that the larger values of X correspond to the smaller values of Y.

d) $cov(A,B) = E[AB] - E[A]E[B]$

$E[AB] = E[(2X+Y)(X-2Y)] = E[2X^2 + YX - 4YX - 2Y^2] = 2E[X^2] - 3E[YX] - 2E[Y^2]$

$E[XY] = E[X]E[Y] = \lambda\left(\frac{a+b}{2}\right)$ X, Y independent

$E[AB] = 2(\lambda^2+\lambda) - 3\lambda\left(\frac{a+b}{2}\right) - 2\left(\frac{1}{b-a}\right)\left(\frac{b^3-a^3}{3}\right)$

$$\boxed{cov(A,B) = 2(\lambda^2+\lambda) - 3\lambda\left(\frac{a+b}{2}\right) - 2\left(\frac{1}{b-a}\right)\left(\frac{b^3-a^3}{3}\right) - \left[2\lambda\left(\frac{a+b}{2}\right)\right]\left[\lambda - a - b\right]}$$

e) Given random variables X and Y, they are independent <u>iff</u> $P(X,Y) = P(X)P(Y)$

independent <u>iff</u> $cov(X,Y) = 0$

$cov(A,B) \neq 0$, so A and B are not independent.

or

$cov(A,B) = cov(2X+Y, X-2Y) = 2cov(X,X) - 4cov(X,Y) + cov(Y,X) - 2cov(Y,Y)$

$= 2 \cdot var(X) - 4(0) + 0 - 2 var(Y)$

$= 2\lambda - 2\frac{(b-a)^2}{12} = \boxed{2\lambda - \frac{(b-a)^2}{6}}$

2. a) $\begin{bmatrix} \sigma_x & \rho\sqrt{\sigma_x \sigma_y} \\ \rho\sqrt{\sigma_x \sigma_y} & \sigma_y \end{bmatrix}$

b) $N(0,1)$

c) $\mu_{y|x} = \mu_y + \frac{\sigma_{xy}}{\sigma_y}(x - \mu_x)$

$= 0 + \frac{(0.25 \cdot 1 \cdot 1)}{1}(x - 0)$

$\boxed{= +0.25x}$

$\sigma_{y|x} = \sigma_y - \frac{\sigma_{xy}\sigma_{yx}}{\sigma_x} = 1 - \frac{(0.25)^2}{1} = \boxed{0.9375}$

$\boxed{N(0.25x, 0.9375)}$

3. a) ML $\underset{\Theta}{\operatorname{argmax}} P(X|\Theta)$ maximize probability of getting x given underlying distribution

MAP $\underset{\Theta}{\operatorname{argmax}} P(\Theta|X)$ fit distribution to data

b) $MAP(\Theta) = \underset{\Theta}{\operatorname{argmax}} P(\Theta|X) = 1$

c) $$MMSE(\Theta) = E[\Theta|X=x] = \int_0^1 \Theta\cdot(0.25\Theta)\, d\Theta + \int_1^8 \Theta\left(\frac{-0.25}{7}(\Theta-8)\right) d\Theta$$

$$= \frac{\Theta^3}{3}\left(\frac{1}{4}\right)\Big|_0^1 + \frac{\Theta^3}{3}\left(\frac{-0.25}{7}\right) - \frac{\Theta^2}{2}\left(\frac{-0.25\cdot 8}{7}\right)\Big|_1^8$$

$$= \frac{1}{12} + \frac{8^3}{21}(-0.25) + \frac{8^2}{7} - \frac{1}{21}(-0.25) - \frac{1}{7}$$

$$= \boxed{3}$$

d) $MLE(\Theta) = \underset{\Theta}{\operatorname{argmax}} P(X|\Theta)$

$= \underset{\Theta}{\operatorname{argmax}} \frac{P(\Theta|X)P(X)}{P(\Theta)}$ Baye's Rule

$= \underset{\Theta}{\operatorname{argmax}} \frac{P(\Theta|X)}{P(\Theta)}$ remove terms that don't depend on Θ

$= \underset{\Theta}{\operatorname{argmax}} P(\Theta|X)$ if $P(\Theta)$ is uniform over all Θ, it can be removed

$= MAP(\Theta)$

$= \boxed{1}$

4. a)

$$MLE(\theta) = \underset{\theta}{\operatorname{argmax}}\ P(X_1 \ldots X_n | \theta)$$

$$= \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^{n} P(x_i|\theta)$$ $x_1 \ldots x_n$ independent

$$= \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^{n} \frac{2x_i}{\theta^2} e^{-\frac{x_i^2}{\theta^2}}$$ for $x_i \geq 0 \ \forall x_i \in \{x_1 \ldots x_n\}$

$$= \underset{\theta}{\operatorname{argmax}} \log\left(\prod_{i=1}^{n}\left(\frac{2x_i}{\theta^2} e^{\frac{-x_i^2}{\theta^2}}\right)\right)$$ log is monotonically increasing

$$= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n}\left(\log(2x_i) - \log\theta^2 + \log(e)^{\frac{-x_i^2}{\theta^2}}\right)$$ log product rule

$$= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n}\left(-\log\theta^2 + \log(e)^{\frac{-x_i^2}{\theta^2}}\right)$$ drop terms which are not a function of θ

$$= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n}\left(-\log\theta^2 - \frac{x_i^2}{\theta^2}\right)$$ $\log(e^x) = x$

$$= \underset{\theta}{\operatorname{argmax}}\ -2n\log\theta - \frac{1}{\theta^2}\sum_{i=1}^{n} x_i^2$$ simplify summation

~ take derivative and set to zero ~

$$-\frac{2n}{\theta} - \left(\sum_{i=1}^{n} x_i^2\right)\left(\frac{-2}{\theta^3}\right) = 0$$

$$2n = \sum_{i=1}^{n} x_i^2 \left(\frac{2}{\theta^2}\right)$$

$$\theta^2 = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2\right)$$

$$\boxed{\hat{\theta}_{ML} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}}$$

b) $$MLE(\theta) = \underset{\theta}{\operatorname{argmax}}\ P(x_1 \ldots x_n \mid \theta)$$

$$= \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^{n} P(x_i \mid \theta)$$ $x_1 \ldots x_n$ are independent

$$= \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^{n} \mathcal{N}(\theta, 1)$$

$$= \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi \cdot 1}} e^{-\frac{(x_i-\theta)^2}{2 \cdot 1}}$$ $\forall x_i \in \{x_1, \ldots x_n\}$

$$= \underset{\theta}{\operatorname{argmax}} \log\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{\frac{-(x_i-\theta)^2}{2}}\right)$$ log transformation

$$= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n}\left[\log(1) - \log(\sqrt{2\pi}) + \log(e)^{\frac{-(x_i-\theta)^2}{2}}\right]$$ log sum rule

$$= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} \log(e)^{\frac{-(x_i-\theta)^2}{2}}$$ remove terms which don't depend on θ

$$= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} -\frac{(x_i-\theta)^2}{2}$$ $\log(e)^x = x$

$$= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} \frac{-(x_i^2 - 2\theta x + \theta^2)}{2}$$ expand the square

$$= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} \frac{-x_i^2}{2} - \theta x_i + \frac{\theta^2}{2}$$ distribute denominator

$$= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} -\theta x + \sum_{i=1}^{n} \frac{\theta^2}{2}$$ remove terms which don't depend on θ and distribute sum

$$= \underset{\theta}{\operatorname{argmax}} \frac{n\theta^2}{2} - \theta \sum_{i=1}^{n} x_i$$ simplify sum

~ take derivative and set to 0 ~

$$\frac{n}{2}(2\theta) - \sum_{i=1}^{n} x_i = 0$$

$$n\theta = \sum_{i=1}^{n} x_i$$

$$\boxed{\hat{\theta}_{MLE} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

intuition: the parameter θ, or the mean of the normal can be estimated by taking an empirical average.

c) $MLE(\theta) = \underset{\theta}{\operatorname{argmax}}\ P(x_1 \ldots x_n \mid \theta)$

$= \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^{n} P(x_i \mid \theta)$ $\quad x_1 \ldots x_n$ are independent

$= \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!}$ $\quad x_i \geq 0 \ \forall x_i \in \{x_1 \ldots x_n\}$

$= \underset{\theta}{\operatorname{argmax}} \log\left(\prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!}\right)$ $\quad$ log transformation

$= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n}\left[\log(\theta)^{x_i} + \log(e)^{-\theta} - \log(x_i!)\right]$ $\quad$ log sum rule

$= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n}\left[\log(\theta)^{x_i} + \log(e)^{-\theta}\right]$ $\quad$ remove terms which don't depend on θ

$= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n}\left(x_i \log(\theta) - \theta\right)$ $\quad \log e^x = x$ and $\log x_1^{x_2} = x_2 \log x_i$

$= \underset{\theta}{\operatorname{argmax}} \log\theta \sum_{i=1}^{n}(x_i) - n\theta$ $\quad$ simplify summation

~ take derivative and set to 0 ~

$$0 = \left(\frac{1}{\theta}\right)\sum_{i=1}^{n} x_i - n$$

$$n = \frac{\sum_{i=1}^{n} x_i}{\theta}$$

$$\boxed{\hat{\theta}_{MLE} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

5. $$\text{MAP}(\theta) = \underset{\theta}{\operatorname{argmax}}\ P(\theta \mid x_1 \ldots x_n, y_1 \ldots y_n)$$

$$= \underset{\theta}{\operatorname{argmax}} \frac{P(x_1 \ldots x_n, y_1 \ldots y_n \mid \theta) P(\theta)}{P(x_1 \ldots x_n, y_1 \ldots y_n)}$$ Baye's Rule

$$= \underset{\theta}{\operatorname{argmax}}\ P(x_1 \ldots x_n, y_1 \ldots y_n \mid \theta) P(\theta)$$ remove terms which don't depend on θ

$$= \underset{\theta}{\operatorname{argmax}}\ P(\theta) \prod_{j=1}^{n} P(x_j, y_j \mid \theta)$$ x_j, y_j independent $\forall (x_j, y_j)$

$$= \underset{\theta}{\operatorname{argmax}} \log\left(P(\theta) \prod_{i=1}^{n} P(x_j, y_j \mid \theta)\right)$$ log transformation

$$= \underset{\theta}{\operatorname{argmax}} \log P(\theta) + \sum_{i=1}^{n} \log P(x_j, y_j \mid \theta)$$ log sum rule

$$= \underset{\theta}{\operatorname{argmax}} \frac{1}{n} \log P(\theta) + \frac{1}{n} \sum_{i=1}^{n} \log P(x_j, y_j \mid \theta)$$ $\underset{x}{\operatorname{argmax}}(x) = \underset{x}{\operatorname{argmax}}(cx)$ where c is a constant

$$= \underset{\theta}{\operatorname{argmin}} -\frac{1}{n} \log P(\theta) - \frac{1}{n} \sum_{i=1}^{n} \log P(x_j, y_j \mid \theta)$$ $\underset{x}{\operatorname{argmin}} f(x) = -\underset{x}{\operatorname{argmax}} f(x)$

$$\boxed{= \underset{\theta}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \ln\left(P(x_j, y_j \mid \theta)\right) - \frac{1}{n} \ln\left(P(\theta)\right)}$$