Problem 1:

(a)

$$P(x_1, x_2 \cdots x_N) = \frac{\lambda^{\sum_{n=1}^{N} x_i}}{\prod_{i=1}^{N} (x_i!)} e^{-N\lambda}$$

(b)

$$\lambda_{ML} = \arg\max_{\lambda} P(x_1, x_2 \cdots x_N)$$

$$= \arg\max_{\lambda} \ln P(x_1, x_2 \cdots x_n)$$

$$= \arg\max_{\lambda} \left(-N\lambda + \sum_{n=1}^{N} x_i \ln\lambda - \sum_{i=1}^{N} \ln(x_i!)\right)$$

to solve λ_M, find

$$\nabla_\lambda -N\lambda + \sum_{i=1}^{N} x_i \ln\lambda - \sum_{i=1}^{N} \ln(x_i!) = 0$$

$$-N + \sum_{i=1}^{N} x_i \times \frac{1}{\lambda} = 0$$

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

(c)

$$\lambda_{MAP} = \arg\max_{\lambda} \ln p(\lambda | X)$$

$$= \arg\max_{\lambda} \ln \frac{p(X|\lambda) p(\lambda | a, b)}{\int_0^{+\infty} p(x_1, x_2 \cdots x_n | \lambda) p(\lambda) d\lambda}$$

$$= \arg\max_{\lambda} \ln p(X|\lambda) + \ln p(\lambda | a, b)$$

$$\nabla_\lambda \; -N\lambda + \sum_{i=1}^{N} x_i \ln\lambda - \sum_{i=1}^{N} \ln(X_i!) + \ln \frac{b^a \lambda^{a-1} e^{-b\lambda}}{\Gamma(a)} = 0$$

$$-N + \frac{\sum_{i=1}^{N} x_i}{\lambda_{MAP}} + (a-1)\frac{1}{\lambda_{MAP}} - b = 0$$

$$b + N = \frac{1}{\lambda_{MAP}} \left(a - 1 + \sum_{i=1}^{N} x_i \right)$$

$$\lambda_{MAP} = \frac{a - 1 + \sum_{i=1}^{N} x_i}{b + N}$$

(d)

$$P(\lambda \mid X) = \frac{P(x_1, \cdots, x_n \mid \lambda) P(\lambda)}{\int_0^{+\infty} P(x_1, x_2 \cdots x_n \mid \lambda) P(\lambda) d\lambda}$$

$$P(\lambda \mid X) \propto P(x_1, \cdots, x_n \mid \lambda) P(\lambda)$$

$$\propto \frac{\lambda^{\sum_{i=1}^{N} x_i}}{\prod_{i=1}^{N} (x_i!)} e^{-N\lambda} \times \frac{b^a \lambda^{a-1} e^{-b\lambda}}{\Gamma(a)}$$

$$\propto \lambda^{(a-1+\sum_{i=1}^{N} x_i)} e^{-(b+N)\lambda}$$

We can recognize from the format.

$$P(\lambda \mid x_1, x_2, \cdots x_n) = \text{gamma}\left(\sum_{i=1}^{N} x_i + a, b+N\right)$$

(e)

$$\mathbb{E}(\lambda \mid x_1, x_2 \cdots x_n)$$

$$= \int_0^{+\infty} \frac{\lambda (b+N)^{\sum_{i=1}^{N} x_i + a} \lambda^{\sum_{i=1}^{N} x_i + a - 1} e^{-(b+N)\lambda}}{\Gamma(\sum_{i=1}^{N} x_i + a)} d\lambda$$

$$= \frac{(b+N)^{\sum_{i=1}^{N} x_i + a}}{\Gamma(\sum_{i=1}^{N} x_i + a)} \int_0^{+\infty} \lambda^{\sum_{i=1}^{N} x_i + a} e^{-(b+N)\lambda} d\lambda$$

$$= \frac{(b+N)^{\sum_{i=1}^{N} x_i + a}}{\Gamma(\sum_{i=1}^{N} x_i + a)} (b+N)^{-(\sum_{i=1}^{N} x_i + a + 1)} \Gamma(\sum_{i=1}^{N} x_i + a + 1)$$

$$= \frac{\sum_{i=1}^{N} x_i + a}{(b+N)} \Rightarrow \text{mean of } \lambda.$$

$$\mathbb{E}(\lambda^2 | x_1, x_2 \cdots x_n)$$

$$= \int_0^{+\infty} \frac{\lambda^2 (b+N)^{\sum_{i=1}^{N} x_i + a} \lambda^{\sum_{i=1}^{N} x_i + a - 1} e^{-(b+N)\lambda}}{\Gamma(\sum_{i=1}^{N} x_i + a)}$$

$$= \frac{(\sum_{i=1}^{N} x_i + a)(\sum_{i=1}^{N} x_i + a + 1)}{(b+N)^2}$$

$$Var(\lambda | x_1, x_2, x_3 \cdots x_n) = \mathbb{E}(\lambda^2 | x_1, x_2 \cdots x_n) - \mathbb{E}^2(\lambda)$$

$$= \frac{(\sum_{i=1}^{N} x_i + a)(\sum_{i=1}^{N} x_i + a + 1) - (\sum_{i=1}^{N} x_i + a)^2}{(b+N)^2}$$

$$= \frac{\sum_{i=1}^{N} x_i + a}{(b+N)^2}$$

λ_{ML} calculates the mode of λ based solely on the data given. λ_{MAP} calculates the mode of λ based on the posterior distribution of λ considering prior distribution of λ and X. If the prior distribution is set to $P(\lambda) = 1$ ($a=1, b=0$ in gamma distribution), λ_{ML} will be the same as λ_{MAP}. $E(\lambda | X)$ calculates the expectation (mean) of posterior distribution. If mode and mean are equal for posterior distribution (e.g, when prior distribution set to normal distribution), they should be the same.

Problem 2:

$$w_{RR} = (\lambda I + X^T X)^{-1} X^T y.$$

$$\mathbb{E}(w_{RR}) = \mathbb{E}\left[(\lambda I + X^T X)^{-1} X^T y\right]$$

$$= \int (\lambda I + X^T X)^{-1} X^T y \, dy.$$

$$= (\lambda I + X^T X)^{-1} X^T \mathbb{E}(y) \Rightarrow \text{given.}$$

$$= (\lambda I + X^T X)^{-1} X^T X w$$

$$Var(y) = \mathbb{E}[(y-\mu)(y-\mu)^T]$$

$$= \mathbb{E}[yy^T - y\mu^T - \mu y^T + \mu\mu^T]$$

$$= \mathbb{E}[yy^T] - \mu\mu^T$$

$$\mathbb{E}[yy^T] = Var(y) + \mu\mu^T$$

$$= \sigma^2 I + X w w^T X^T$$

$$Var(W_{RR}) = \mathbb{E}\left[(W_{RR} - \mathbb{E}(W_{RR}))(W_{RR} - \mathbb{E}(W_{RR}))^T\right]$$

$$= \mathbb{E}[W_{RR} W_{RR}^T] - \mathbb{E}[W_{RR}]\mathbb{E}[W_{RR}]^T.$$

$$= \mathbb{E}\left((\lambda I + X^T X)^{-1} X^T y y^T X ((\lambda I + X^T X)^{-1})^T\right) -$$

$$\mathbb{E}[W_{RR}]\mathbb{E}[W_{RR}]^T$$

$$= (\lambda I + X^T X)^{-1} X^T \mathbb{E}(y y^T) X ((\lambda I + X^T X)^{-1})^T -$$

$$\mathbb{E}[W_{RR}]\mathbb{E}[W_{RR}]^T$$

$$= (\lambda I + X^T X)^{-1} X^T (\sigma^2 I + X w w^T X^T) X ((\lambda I + X^T X)^{-1})^T -$$

$$\mathbb{E}(W_{RR})\mathbb{E}[W_{RR}]^T$$

Let $Z = (I + \lambda (X^T X)^{-1})^{-1}$

$$Var(W_{RR}) = Z (X^T X)^{-1} X^T (\sigma^2 I + X w w^T X^T) X \cdots$$

$$(Z (X^T X)^{-1})^T - Z w w^T Z^T$$

$$Var(W_{RR}) = \sigma^2 Z (X^T X)^{-1} Z^T + Z w w^T Z^T - Z w w^T Z^T$$

$$= \sigma^2 Z (X^T X)^{-1} Z^T$$