(d) when $\gamma = 1-\lambda$, we have $p(x,y) = \lambda(1-\lambda) e^{-\lambda x - (1-\lambda) y}$

$$l(\lambda) = \ln \prod_{i=1}^{N} \lambda(1-\lambda) e^{-\lambda x_i - (1-\lambda) y_i} = \sum_{i=1}^{N} \ln[\lambda(1-\lambda)] - \sum_{i=1}^{N} [\lambda x_i + (1-\lambda) y_i]$$

$$= N \ln[\lambda(1-\lambda)] - N\lambda\bar{x} - N(1-\lambda)\bar{y}$$

$$\frac{\partial l(\lambda)}{\partial \lambda} = \frac{N}{\lambda} - \frac{N}{1-\lambda} - N\bar{x} + N\bar{y} = 0$$

$$\Rightarrow \quad (\bar{x}-\bar{y})\lambda^2 - (\bar{x}-\bar{y}+2)\lambda + 1 = 0$$

$$\hat{\lambda} = \frac{(\bar{x}-\bar{y}+2) + \sqrt{(\bar{x}-\bar{y})^2 + 4}}{2(\bar{x}-\bar{y})}$$

Exercise 2:

(a). $\hat{\beta} = (X^T X)^{-1} X^T \mathbf{y}$ where $\mathbf{y} = \mathbf{X}\beta + \varepsilon$

$$\therefore \quad \hat{\beta} = (X^T X)^{-1} X^T (X\beta + \varepsilon) = \beta + (X^T X)^{-1} X^T \varepsilon$$

since $\varepsilon \sim N(0, \sigma^2)$

$\hat{\beta}$ also follows a Gaussian distribution, with

$$E(\hat{\beta}) = E(\beta + (X^T X)^{-1} X^T \varepsilon) = \beta + (X^T X)^{-1} X^T E(\varepsilon)$$

$$= \beta + 0 = \beta$$

$$Var(\hat{\beta}) = E[\hat{\beta} - E(\hat{\beta})][\hat{\beta} - E(\hat{\beta})]^T$$

$$= E[\hat{\beta} - \beta][\hat{\beta} - \beta]^T$$

$$= E[(X^T X)^{-1} X^T \varepsilon \varepsilon^T X (X^T X)^{-1}]$$

$$= (X^T X)^{-1} X^T E(\varepsilon\varepsilon^T) X (X^T X)^{-1}$$

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

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

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