1. (a)

	H	C
PH	882	10
PC	18	90

Given: $p(H) = \theta$; $p(C) = 1-\theta$; $p(PH|H) = p$;

$p(PC|H) = 1-p$; $p(PH|C) = 1-q$; $p(PC|C) = q$

The joint likelihood of the data is:

$$L(p, q, \theta) = p(H, PH)^{882} \, p(H, PC)^{18} \, p(C, PH)^{10} \, p(C, PC)^{90}$$

Now $p(H, PH) = p(H) . p(PH|H) = \theta p$

Similarly for the other 3 cases...

$$\Rightarrow L(p, q, \theta) = (\theta p)^{882} (\theta(1-p))^{18} ((1-\theta)(1-q))^{10} ((1-\theta) q)^{90}$$

$$= \theta^{900} (1-\theta)^{100} p^{882} (1-p)^{18} q^{90} (1-q)^{10}$$

$$\frac{\partial L}{\partial \theta} = 900 \, \theta^{899} (1-\theta)^{100} p^{882} \ldots$$

$$\cancel{+} - \theta^{900} . 100 (1-\theta)^{99} p^{882} \ldots$$

$$= \{ 900(1-\theta) \cancel{+} - \theta . 100 \} \{ \theta^{899} (1-\theta)^{99} p^{882} \ldots \}$$

For max. likelihood, $\frac{\partial L}{\partial \theta} = 0$

$\theta = 0$, $\theta = 1$ are clearly minima, not maxima

$$\Rightarrow 900(1-\theta_{ML}) \cancel{+} - 100 \, \theta_{ML} = 0$$

$$\Rightarrow \theta_{ML} = \frac{900}{1000} = \boxed{0.9}$$

Similarly, $p_{ML} = \frac{882}{900} = \boxed{0.98}$

$q_{ML} = \frac{90}{100} = \boxed{0.9}$

1.(b) $E(L_{FOB}) = 0 \cdot P(H, PH) + K \cdot P(C, PH) + 1 \cdot P(H, PC) + 0 \cdot P(C, PC)$

$= K(1-\theta)(1-q) + \theta(1-p)$

[using ML estimates] $= K(0.1)(0.1) + 0.9(0.02) = \boxed{0.01K + 0.018}$

1.(c) $E(L_{CTS}) = K(1-\theta)(1-q') + \theta(1-p')$

[θ_{ML} is the same for the same cohort]

$= K(0.1)(0.04) + 0.9(0.2)$

$= \boxed{0.004K + 0.18}$

For critical K, $E(L_{CTS}) = E(L_{FOB})$

$\Rightarrow 0.004K + 0.18 = 0.01K + 0.018$

$0.006K = 0.162$

$K = \boxed{27}$

2. (a) $p(\underline{k} \mid \lambda) = \prod_{i=1}^{N} \frac{\lambda^{k_i} e^{-\lambda}}{k_i!} = \frac{\lambda^{\sum_i k_i} e^{-N\lambda}}{\prod_i k_i!}$

$\log p(\underline{k} \mid \lambda) = \sum_i k_i \log \lambda - N\lambda - \log\left(\prod_i k_i!\right)$

$\frac{\partial \log p(\underline{k} \mid \lambda)}{\partial \lambda} = \frac{\sum_i k_i}{\lambda} - N$

$\Rightarrow \lambda_{ML} = \boxed{\frac{\sum_i k_i}{N}}$

2.(b) $\lambda_{ML} = \dfrac{6+32+5+4+4}{5} = \boxed{10.2}$ minutes

Clearly distorted by outlier 32.

2.(c) $$p(\lambda | \underline{k}) = \frac{p(\underline{k}|\lambda) \cdot p(\lambda)}{p(\underline{k})}$$

$$= \frac{1}{p(\underline{k})} \cdot \frac{\lambda^{\sum_i k_i} e^{-N\lambda}}{\prod_i k_i!} \cdot \frac{1}{K} \lambda^{\alpha-1} e^{-\beta\lambda}$$

$$= \frac{1}{K p(\underline{k})} \cdot \frac{\lambda^{\sum_i k_i + \alpha - 1} e^{-(N+\beta)\lambda}}{\prod_i k_i!}$$

By analogy with part (a),

$$\lambda_{MAP} = \boxed{\frac{\sum_i k_i + \alpha - 1}{N+\beta}}$$

2.(d) We see that prior has effect of adding on β additional 'observations', having a total value of $\alpha - 1$, to the actual data. Thus, if $\beta = 10$, and our prior expectation of the average interval is 5, then $\boxed{\alpha = 51}$

$$\Rightarrow \lambda_{MAP} = \frac{51+51-1}{5+10} = \frac{101}{15} = \boxed{6.73}$$

Thus, with a small amount of data, we have shifted the estimate closer to the prior expectation and reduced the distorting effect of one outlier.

3. $$p(C_1|\underline{x}) = \frac{p(\underline{x}|C_1)\cdot p(C_1)}{p(\underline{x})}$$

$$= \frac{p(\underline{x}|C_1)\cdot p(C_1)}{p(\underline{x}|C_1)\cdot p(C_1) + p(\underline{x}|C_2)\cdot p(C_2)}$$

$$= \frac{1}{1 + \frac{p(\underline{x}|C_2)\,p(C_2)}{p(\underline{x}|C_1)\,p(C_1)}} = \boxed{\frac{1}{1+e^{-a}}}$$

where $a = \ln \frac{p(\underline{x}|C_1)\,p(C_1)}{p(\underline{x}|C_2)\,p(C_2)}$ [to show this is $\underline{w}^T\underline{x} + w_0$]

$$\Rightarrow a = \ln \left\{ \frac{\cancel{\frac{1}{(2\pi)^{n/2}}\frac{1}{|\Sigma|^{1/2}}} \exp\left\{-\frac{1}{2}(\underline{x}-\mu_1)^T\Sigma^{-1}(\underline{x}-\mu_1)\right\}\cdot\theta}{\cancel{\frac{1}{(2\pi)^{n/2}}\frac{1}{|\Sigma|^{1/2}}} \exp\left\{-\frac{1}{2}(\underline{x}-\mu_2)^T\Sigma^{-1}(\underline{x}-\mu_2)\right\}\cdot(1-\theta)} \right\}$$

$$= -\frac{1}{2}(\underline{x}-\mu_1)^T\Sigma^{-1}(\underline{x}-\mu_1) + \frac{1}{2}(\underline{x}-\mu_2)^T\Sigma^{-1}(\underline{x}-\mu_2) + \ln\left(\frac{\theta}{1-\theta}\right)$$

[quadratic terms cancel]

$$= -\frac{1}{2}\left\{\cancel{\underline{x}^T\Sigma^{-1}\underline{x}} - \underline{x}^T\Sigma^{-1}\mu_1 - \mu_1^T\Sigma^{-1}\underline{x} + \mu_1^T\Sigma^{-1}\mu_1 - \cancel{\underline{x}^T\Sigma^{-1}\underline{x}} + \underline{x}^T\Sigma^{-1}\mu_2 + \mu_2^T\Sigma^{-1}\underline{x} - \mu_2^T\Sigma^{-1}\mu_2\right\} + \ln\left(\frac{\theta}{1-\theta}\right)$$

[Re-writing dot products]

$$= -\frac{1}{2}\left\{-(\Sigma^{-1}\mu_1)^T\underline{x} - \mu_1^T\Sigma^{-1}\underline{x} + (\Sigma^{-1}\mu_2)^T\underline{x} + \mu_2^T\Sigma^{-1}\underline{x}\right\} + \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \ln\left(\frac{\theta}{1-\theta}\right)$$

We know that Σ^{-1} is symmetric : $\Sigma^{-1} = (\Sigma^{-1})^T$

$$\therefore \; (\mu_1^T \Sigma^{-1}) = (\Sigma^{-1} \mu_1)^T$$

$$\Rightarrow a = -\frac{1}{2}\Big\{ -(\Sigma^{-1}\mu_1)^T \underline{x} - (\Sigma^{-1}\mu_1)^T \underline{x} + (\Sigma^{-1}\mu_2)^T \underline{x} + (\Sigma^{-1}\mu_2)^T \underline{x} \Big\} + \underbrace{\frac{1}{2}\mu_2^T \Sigma^{-1} \mu_2 - \frac{1}{2}\mu_1^T \Sigma^{-1}\mu_1 + \ln\left(\frac{\theta}{1-\theta}\right)}_{\triangleq\, w_0}$$

$$= +\frac{1}{\cancel{2}}\Big\{+\cancel{2}\big[\underbrace{\Sigma^{-1}(\mu_1 - \mu_2)}_{\triangleq\, \underline{w}}\big]^T \underline{x}\Big\} + w_0$$

$$= \boxed{\underline{w}^T \underline{x} + w_0}$$