Q1 1) $P(D|\lambda) = \dfrac{e^{-n\lambda} \cdot \lambda^{\Sigma x_i}}{\prod_{i=1} x_i!} \Rightarrow \log P(D|\lambda) = -n\lambda + \Sigma x_i \cdot \log(\lambda) - \log(\prod_{i=1} x_i!)$

2) $\dfrac{d}{d\lambda} \log P(D|\lambda) = -n + \dfrac{\Sigma x_i}{\lambda} - 0$

3) $-n + \dfrac{\Sigma x_i}{\lambda} = 0 \Rightarrow \dfrac{\Sigma x_i}{\lambda} = n \Rightarrow \boxed{\dfrac{\Sigma x_i}{n} = \hat{\lambda}_{MLE}}$

~~Q2~~

Q2 $f(x_i, \lambda) = \dfrac{e^{-N\lambda} \lambda^{\sum_{i=1}^{N} x_i}}{\prod_{i=1}^{N} x_i!} = \dfrac{e^{-\lambda\beta} \lambda^{\alpha-1} \beta^{\alpha}}{\Gamma(\alpha)}$

$$f(x) = \int_0^\infty \frac{\beta^\alpha}{\prod_{i=1}^{N} x_i! \Gamma(\alpha)} = e^{-N\lambda} e^{-\lambda\beta} \lambda^{\sum_{i=1}^{N} x_i} \lambda^{\alpha-1} \, d\lambda$$

$$= \frac{\beta^\alpha}{\prod_{i=1}^{N} x_i! \Gamma(\alpha)} \int_0^\infty e^{-\lambda(N+\beta)} \lambda^{\sum_{i=1}^{N} x_i + \alpha - 1} \, d\lambda$$

$$P(X|D) = \frac{\dfrac{\cancel{\beta^\alpha}}{\prod_{i=1}^{N} x_i! \cancel{\Gamma(\alpha)}} e^{-N\lambda} e^{-\lambda(N+\beta)} \lambda^{\sum_{i=1}^{N} x_i + \lambda - 1}}{\dfrac{\cancel{\beta^\alpha}}{\prod_{i=1}^{N} x_i! \cancel{\Gamma(\alpha)}} \quad \dfrac{\Gamma(\sum_{i=1}^{N} x_i + \alpha)}{(N+\beta)^{\sum_{i=1}^{N} x_i + \alpha}}}$$

$$= \frac{e^{-\lambda(N+\beta)} \lambda^{\sum_{i=1}^{N} x_i + \alpha - 1} (N+\beta)^{\sum_{i=1}^{N} x_i + \alpha}}{\Gamma(\sum_{i=1}^{N} x_i + \alpha)}$$

$$\ln P(\lambda|D) = \frac{d}{d\lambda}(-\lambda(N+\beta)) + \frac{d}{d\lambda}(\ln\lambda(\sum_{i=1}^{N} x_i + \alpha - 1))$$

$$0 = -(N+\beta) + \frac{\sum_{i=1}^{N} x_i + \alpha - 1}{\lambda}$$

$$\frac{\sum_{i=1}^{N} x_i + \alpha - 1}{N+\beta} = \lambda$$

$$\boxed{\hat{\lambda}_{MAP} = \frac{\sum_{i=1}^{N} x_i + \alpha - 1}{N+\beta}}$$

Q3) $f(y_1, y_2, \ldots y_n | \lambda) = \dfrac{e^{-n\lambda} \lambda^{\sum_{i=1}^{n} y_i}}{\prod_{i=1}^{n} y_i!}$ prior: $\dfrac{e^{-\beta\lambda} \lambda^{\alpha-1} \beta^{\alpha}}{\Gamma\alpha}$

$$\pi(\lambda | D) \propto f(y_1, \ldots, y_n | \lambda) \text{ Pr.}$$

$$= \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^{n} y_i}}{\prod_{i=1}^{n} y_i!} \cdot \frac{e^{-\beta\lambda} \lambda^{\alpha-1} \beta^{\alpha}}{\Gamma\alpha}$$

$$= e^{-\lambda(\beta+n)} \lambda^{\sum_{i=1}^{n} y_i + \alpha - 1}$$

$\alpha_p = \sum_{i=1}^{n} y_i + \alpha - 1$ $\quad \beta_p = \beta + n$

therefore gamma distribution is conjugate prior to the Poison. ✓

Q4) When nominal attributes are represented as ordinal #s there is a heiarchy implied by the increasing numbers, so the distance from one point in the class to another in the class will be different then another comparison of points in the class. But if they are represented as an array of binary options, then the distances will alway be the same in the same class because there will always only be one one in differing positions.

$\sqrt{2}$	Private	$\sqrt{2}$	1	0	0
$\sqrt{2}$	State-gov		0	1	0
	Never worked		0	0	1

$1 = \sqrt{1}$	Private	2	1
	State-gov		2
$1 = \sqrt{1}$	Never worked		3

Q5) $1967/8000 \approx 24.6\%$ of training data has income >50K.

↑mistake $1 - 24.6 = 75.4\%$

A 70% accuracy would not be a good model because 75.4% of the class label is >50 K, so if you guessed the same attribute everytime you'd get a 75.4% sucess rate $70 < 75.4$. If the distribution was tighter then 70% would be better, although still not the best.

---

age: 1D education: 1D capital-gains: 1D
capital-loss: 1D hours-per-week: 1D
workclass: 7D marital-status: 7D
occupation: 14D relationship: 6D race: 5D
native-country: 40D sex: 1D

Q6)

$$\|x-z\|_2 = \sqrt{\sum_{i=1}^{d}(x_i-z_i)^2}$$