Question-1:

a) We are asked to show $I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$ where 'I ' is information gain and 'H' is entropy function. We are asked to do this by using '*KL-divergence*'.

$$I(X,Y) = KL(p(x,y)||p(x)p(y))$$

$$= -\sum_x \sum_y p(x,y)\ log_2 \frac{p(x)p(y)}{p(x,y)}$$

Here we need to remember $\sum_x p(x,y) = p(y)$ and $p(x,y) = p(y|x)p(x) or = p(x|y)p(y)$. After expanding logarithm and doing simplifications with the rules I have given above we reach to final equation:

$$-\sum_y p(y) log_2 p(y) + \sum_x p(x) \sum_y p(y|x) log_2 p(y|x)$$

If we remember the formula of entropy, we can see that this equation is:

$$H(Y) - H(Y|X)$$

With little differences and different conditional probability expansion $H(X) - X(X|Y)$ can be shown. Idea is same.

b) We need to find under which conditions $I(X,Y) = 0$. If we look at the expansion of *information gain* function with *KL-divergence I* gave above, if $p(x,y) = 0$ all equation is clearly zero. We need to consider when $p(x,y) = 0$, and we can remember that if probabilities are independent their joint probability is zero. This result satisfies what we need.

Question-2: Question asks for us to drive the formula of H(X), which is entropy, for random variable X that has normal distribution. First it is given that:

$$H(X) = -\int p(x) ln p(x) dx$$

We know the probability density of normal distribution. After some calculations equation becomes:

$$= \frac{1}{2}(ln(2\pi\sigma^2) + 1)$$

Variance of normal distribution can be written as:

$$\sigma^2 = \int p(x)(x - \mu)^2 \, dx$$

From here we can see that if $\sigma^2 < \frac{1}{2\pi e}$ H(x) becomes less than zero. So we observed that entropy for continuous random variables can be negative, unlike entropy for discrete random variables.

Question-3:

a) Let D be the random variable that denotes the existence of disease, and T denotes results of tests. P(D=1) and P(T=1) will be simply referred as P(D) and P(T).

We are asked to calculate $P(T)$ which is $P(T) = P(T, D) + P(T, D')$

$$From conditional probability P(T) = P(T|D)P(D) + P(T|D')P(D')$$

All these values are given at the question. It follows that:

$$P(T) = 0.95 * 0.01 + 0.05 * 0.99 = 0.059$$

b) We are asked to calculate P(D|T), which is probability of infected by the disease given that test is positive. We will use Bayes' Rule for this.

$$P(D|T) = \frac{P(T|D)P(D)}{P(T)}$$

It follows that:

$$P(D|T) = \frac{0.95 * 0.01}{0.059} = 0.16$$

Question-4:

a) We are asked to show that $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the MLE of lambda and it is unbiased ($E(\hat{\lambda}) = \lambda$)

First, we need to find log-likelihood where D denotes random variable X's that are Poisson distributed. After some calculations we find it as:

$$\ln P(D|\lambda) = -n\lambda + \sum_{i=1}^{n} X_i ln\lambda - \ln(X_i!)$$

We know that $MLE\hat{\lambda} = argmax_{\lambda} \ln P\,(D|\lambda)$and can be obtained by taking derivative of $lnP(D|\lambda)$ with respect to lambda end equal it zero. After some computations:

$$\frac{d}{d\lambda} \ln P(D|\lambda) = -n + \frac{1}{\lambda}\sum_{i=1}^{n} X_i = 0$$

From here we can see that $\lambda = \frac{1}{n}\sum_{i=1}^{n} X_i$ in this case it becomes $\hat{\lambda}$.

We can show that $\hat{\lambda}$ is unbiased by taking its expectancy and showing it is equal to lambda.

$E(\hat{\lambda}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$ above we showed that $E(X_i) = \lambda$. Then the equation becomes:

$$= \frac{1}{n}\sum_{i=1}^{n} \lambda \, which \; is = \lambda, hence \; showed.$$

b) We are asked to show $P(\lambda|D)$ which is the posterior distribution over lambda. First, we must remember the Bayes' Rule and apply it for $P(\lambda|D)$. Because the denominator is independent from lambda, we can discard it. Hence, we are left with:

$$P(\lambda|D) \propto \mathrm{P(D|\lambda)p(\lambda)}$$

Where P is probability mass function for Poisson distribution that we calculated at previous part, p is probability density function for Gamma distribution, which is given at the question. After inserting those and discarding elements independent from λ we are left with:

$$P(\lambda|D) \propto \lambda^{(X_i+\alpha-1)} e^{(-n\lambda-\beta\lambda)}$$

From here we can see that posterior distribution has the same distribution as Gamma with $\alpha = \sum_{i=1}^{n} X_i + \alpha$ and $\beta = n + \beta$.

c) We are asked to find $\hat{\lambda}$ such that $\hat{\lambda} = argmax_{\lambda} lnP(\lambda|D)$. We calculated $P(\lambda|D)$ above. After some calculations we are left with:

$$\boldsymbol{lnP(\lambda|D) = (\textstyle\sum_{i=1}^{n} X\_\{i\} + \alpha - 1)\, ln\, \lambda - (n + \beta)\, \lambda}$$

From here, like we did for MLE, we take the derivative of $lnP(\lambda|D)$ and set it to zero. As a result, we are left with:

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} X_i + \alpha - 1}{n + \beta}$$ Which is the result.