(1) (a) ML Estimator for gaussian distribution:-

$$\hat{\mu} = \frac{\Sigma x_i}{n}$$ (proved in class)

(b) Prior distribution = Gaussian.

Using result proved in class:-

$$\hat{\mu} = \frac{\bar{x}\sigma_0^2 + \mu_0 \frac{\sigma^2}{N}}{\sigma_0^2 + \frac{\sigma^2}{N}}$$

where $\bar{x}$ = sample mean.

σ_0^2 = variance of prior distribution.

μ_0 = mean of prior distribution.

N = no. of data points.

σ^2 = true variance of the sample.

(c).

Part - 2 (MAP)

$$M = \text{mean of the Gaussian}$$

$$P_M(m) = \frac{1}{11.5 - 9.5} = \frac{1}{2}$$

$$P(x_i|M) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i - m)^2}{2\sigma^2}}$$

$$P(\{x_i\}_{i=1}^{n}, M) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum(x_i - m)^2}$$

$$P(\{x_i\}_{i=1}^{n}) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \int_{9.5}^{11.5} e^{-\frac{1}{2\sigma^2}\sum(x_i - m)^2} \times \frac{1}{2}\, dm.$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \times \frac{1}{2} \int_{9.5}^{11.5} e^{-\frac{1}{2\sigma^2}\left(\sum x_i^2 + nm^2 - 2m\sum x_i\right)} dm$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \times \frac{1}{2} \cdot e^{\frac{\sum x_i^2}{2\sigma^2}} \int_{9.5}^{11.5} e^{-\frac{1}{2\sigma^2}\left(nm^2 - 2m\sum x_i\right)} dm.$$

$$P(M|x) = \frac{P(x|m) \times P(m)}{P(x)}$$

$$= \frac{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\Sigma(x_i - m)^2} \times \frac{1}{2}}{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \times \frac{1}{2}\, e^{\frac{\Sigma x_i}{2\sigma^2}} \int_{9.5}^{11.5} e^{-\frac{1}{2\sigma^2}(nm^2 - 2m\Sigma x_i)}\, dm}$$

$$= \frac{e^{-\frac{1}{2\sigma^2}(nm^2 - 2m\Sigma x_i)}}{\int_{9.5}^{11.5} e^{-\frac{1}{2\sigma^2}(nm^2 - 2m\Sigma x_i)}\, dm}$$

~~Denominator is a term in~~

Denominator is independent of m.

So to find mode, i.e. the value of M at which $P(M|x)$ is maximum we have to maximise numerator.

$$e^{-\frac{1}{2\sigma^2}(nm^2 - 2m\sum x_i)} = g(m)$$

~~diff this to be~~ maximum,

on

$$\frac{dg(m)}{dm} = 2mn - 2\sum x_i = 0$$

$$\Rightarrow \boxed{m = \frac{\sum x_i}{n}}$$

Interpretation:-

① As N increases, the relative error decrease for all the three estimators.

② The estimator ~~oth~~ with prior distribution = Gaussian (mean = 10.5, variance = 1) is preffered, as the relative error is less than the other estimator by a factor of ~~a~~ 10 (approx).