Q1:

Part A)

$p(x) = \mu^x(1-\mu)^{1-x} \quad , \quad x \in \{0,1\}$

Likelihood: $\prod_{i=1}^{n} \mu^{x_i}(1-\mu)^{1-x_i} = \mu^{\sum_{i=1}^{n} x_i}(1-\mu)^{\sum_{i=1}^{n}(1-x_i)} = \mu^{\sum_{i=1}^{n} x_i}(1-\mu)^{(n-\sum_{i=1}^{n} x_i)}$

Log-likelihood: $(\sum_{i=1}^{n} x_i)\ln(\mu) + (n - \sum_{i=1}^{n} x_i)\, ln(1-\mu)$

MLE: $\frac{\sum_{i=1}^{n} x_i}{\mu} - \frac{n-\sum_{i=1}^{n} x_i}{1-\mu} = 0 \quad \rightarrow \quad \sum_{i=1}^{n} x_i - \mu\sum_{i=1}^{n} x_i - \mu n + \mu\sum_{i=1}^{n} x_i = 0 \rightarrow \quad \mu = \frac{\sum_{i=1}^{n} x_i}{n}$

Part B)

For this question: $\mu = \frac{2}{10} = 0.2$

Part C)

Considering uniform distribution for the parameter makes MAP estimation equivalent to MLE. Not having μ in the distribution formula, the prior will be removed after taking derivation.

Part D)

MAP with Gaussian prior: $\frac{1}{\sqrt{2\pi\frac{1}{2}}}\exp\left(-\frac{\left(\mu-\frac{1}{2}\right)^2}{2\left(\frac{1}{4}\right)}\right)\prod_{i=1}^{n} \mu^{x_i}(1-\mu)^{1-x_i} = \frac{1}{\sqrt{\frac{\pi}{2}}}\exp\left(-\frac{\left(\mu-\frac{1}{2}\right)^2}{\frac{1}{2}}\right)\mu^{\sum_{i=1}^{n} x_i}(1-\mu)^{(n-\sum_{i=1}^{n} x_i)}$

After using natural logarithm: $ln\frac{\sqrt{2}}{\sqrt{\pi}} - \frac{\left(\mu-\frac{1}{2}\right)^2}{\frac{1}{2}} + (\sum_{i=1}^{n} x_i)\ln(\mu) + (n - \sum_{i=1}^{n} x_i)\, ln(1-\mu) = A$

$$\frac{\partial A}{\partial \mu} = -4\left(\mu - \frac{1}{2}\right) + \frac{\sum_{i=1}^{n} x_i}{\mu} - \frac{n - \sum_{i=1}^{n} x_i}{1-\mu} = -4\mu + 2 + \frac{\sum_{i=1}^{n} x_i}{\mu} - \frac{n - \sum_{i=1}^{n} x_i}{1-\mu} = 0$$

$$\rightarrow \quad -4\mu^2 + 4\mu^3 + 2\mu - 2\mu^2 + \sum_{i=1}^{n} x_i - \mu\sum_{i=1}^{n} x_i - \mu n + \mu\sum_{i=1}^{n} x_i = 0$$

$$\rightarrow \quad 4\mu^3 - 6\mu^2 + (2-n)\mu + \sum_{i=1}^{n} x_i = 0$$

μ is the roots of $\mu^3 - 1.5\mu^2 + \left(\frac{1}{2} - \frac{n}{4}\right)\mu + \frac{\sum x_i}{4} = 0$

For this question: $\mu = -1, 0.25$ and 2.25. Just $\mu = 0.25$ is acceptable as it should be in the range $[0,1]$.

Part E)

We see the MAP estimation with Gaussian prior is better than MLE since it is closer to 0.5 (if we have a fair coin).

Q2:

$x = [x_1, \dots, x_n]$ $\qquad y(x_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_i - T)^2}{2{\sigma_i}^2}\right), \qquad i = 1, \dots, n$

Part A:

$$Likelihood = \prod_{i=1}^{n} y(x_i) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(x_i - t)^2}{2{\sigma_i}^2}\right)\right) = 2\pi^{-\frac{n}{2}} \times \prod_{i=1}^{n} {\sigma_i}^{-1} \times \exp\left(\sum_{i=1}^{n} \left(-\frac{(x_i - t)^2}{2{\sigma_i}^2}\right)\right)$$

Part B:

$$\ln(Likelihood) = -\frac{n}{2}\ln(2\pi) + \ln\left(\sum_{i=1}^{n} {\sigma_i}^{-1}\right) + \sum_{i=1}^{n} \left(-\frac{(x_i - t)^2}{2{\sigma_i}^2}\right)$$

$$MLE = \frac{\partial \ln(Likelihood)}{\partial t} = \sum_{i=1}^{n} (-2)\left(-\frac{(x_i - t)}{2{\sigma_i}^2}\right) = 0 \quad \rightarrow \quad \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} t = nt \quad \rightarrow \quad t = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Q3:

$$likelihood = \prod_{k=1}^{N} p(x_k, \theta) = \prod_{k=1}^{N} \theta^2 x_k \exp(-\theta x_k) U(x_k) = \theta^{2N}\left(\prod_{k=1}^{N} x_k\right) \exp\left(-\theta \sum_{k=1}^{N} x_k\right)\left(\prod_{k=1}^{N} U(x_k)\right)$$

$$log - likelihood = 2Nlog(\theta) + \sum_{k=1}^{N} \log(x_k) - \theta \sum_{k=1}^{N} x_k + \sum_{k=1}^{N} \log(U(x_k))$$

$$\frac{\partial(log - likelihood)}{\partial \theta} = \frac{2N}{\theta} - \sum_{k=1}^{N} x_k = 0 \quad \rightarrow \quad \theta = \frac{2N}{\sum_{k=1}^{N} x_k}$$

Q4:

$$MAP: \left(\frac{\mu}{\sigma_\mu^2} \exp\left(-\frac{\mu^2}{2\sigma_\mu^2}\right)\right) \prod_{i=1}^{N} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2}(x_i - \mu)^2\right)\right) = \frac{\mu}{\sigma_\mu^2} \exp\left(-\frac{\mu^2}{2\sigma_\mu^2}\right)\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^{N}(x_i - \mu)^2\right)$$

Applying log: $A = \ln\left(\frac{\mu}{\sigma_\mu^2}\right) - \frac{\mu^2}{2\sigma_\mu^2} + Nln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i - \mu)^2$

$$\frac{\partial A}{\partial \mu} = \frac{1}{\mu} - \frac{\mu}{\sigma_\mu^2} + \frac{1}{\sigma^2}\sum_{i=1}^{N} x_i - \frac{N\mu}{\sigma^2} = \frac{1}{\mu} + Z - \mu\left(\frac{1}{\sigma_\mu^2} - \frac{N}{\sigma^2}\right) = \frac{1}{\mu} + Z - \mu R = 0 \rightarrow \quad 1 - R\mu^2 + Z\mu = 0$$

$\mu = \frac{-Z \pm \sqrt{Z^2 + 4R}}{-2R} = \frac{Z}{2R}\left(1 \mp \sqrt{1 + \frac{4R}{Z^2}}\right)$ In Rayleigh distribution of μ, only positive values are acceptable for it so:

$$\mu = \frac{Z}{2R}\left(1 + \sqrt{1 + \frac{4R}{Z^2}}\right)$$

Q5:

$$Q(\theta,\theta(t)) = \sum_{k=1}^{N}\sum_{j=1}^{J} p(j|x_k,\theta(t))(-\frac{l}{2}\ln(\sigma_j^2) - \frac{\|x_k-\mu_j\|^2}{2\sigma_j^2} + \ln(p_j)) - \lambda(\sum_{j=1}^{J} p_j - 1)$$

$$\frac{\partial Q(\theta,\theta(t))}{\partial \mu_j} = \sum_{k=1}^{N} p(j|x_k,\theta(t))(-\frac{(-2)}{2\sigma_j^2}\|x_k-\mu_j\|) = 0 \;\rightarrow\; \sum_{k=1}^{N} p(j|x_k,\theta(t))x_k = \sum_{k=1}^{N} p(j|x_k,\theta(t))\,\mu_j$$

$$\mu_{j+1} = \frac{\sum_{k=1}^{N} p(j|x_k,\theta(t))x_k}{\sum_{k=1}^{N} p(j|x_k,\theta(t))}$$: Formula 2.98

$$\frac{\partial Q(\theta,\theta(t))}{\partial \sigma_j^2} = \sum_{k=1}^{N} p(j|x_k,\theta(t))\left(\frac{-l}{2\sigma_j^2} + \frac{2\|x_k-\mu_j\|^2}{4\sigma_j^4}\right) = 0 \rightarrow \; \sigma_{j+1}^2 = \frac{\sum_{k=1}^{N} p(j|x_k,\theta(t))\|x_k-\mu_j\|^2}{-l\sum_{k=1}^{N} p(j|x_k,\theta(t))}$$: Formula 2.99

Lagrange: $A = Q(\theta,\theta(t)) - \lambda(\sum_{j=1}^{J} p_j - 1)$

$$\frac{\partial A}{\partial p_j} = \sum_{k=1}^{N} p(j|x_k,\theta(t))\frac{1}{p_j} - \lambda = 0 \;\rightarrow\; p_{j+1} = \frac{1}{\lambda}\sum_{k=1}^{N} p(j|x_k,\theta(t))$$

We know: $\sum_{j=1}^{N} p_{j+1} = 1 \;\rightarrow\; \sum_{j=1}^{N}\frac{1}{\lambda}\sum_{k=1}^{N} p(j|x_k,\theta(t)) = \sum_{k=1}^{N}\frac{1}{\lambda}\sum_{j=1}^{N} p(j|x_k,\theta(t)) = \frac{1}{\lambda}\sum_{k=1}^{N}\sum_{j=1}^{N} p(j|x_k,\theta(t)) = \frac{1}{\lambda}\sum_{k=1}^{N} 1 = \frac{N}{\lambda} = 1 \;\rightarrow\; N = \lambda \;\rightarrow\; p_{j+1} = \frac{1}{N}\sum_{k=1}^{N} p(j|x_k,\theta(t))$: Formula 2.100

Q6:

$$likelihood(p) = \prod_{i=1}^{m} p(k_i) = \prod_{i=1}^{m}\frac{N!}{k_i!(N-k_i)!}p^{k_i}(1-p)^{N-k_i} = \frac{N!^m}{\prod_{i=1}^{m} k_i!(N-k_i)!}p^{\sum k_i}(1-p)^{mN-\sum k_i}$$

$$log-likelihood(p) = \log\left(\frac{N!^m}{\prod_{i=1}^{m} k_i!(N-k_i)!}\right) + \sum_{i=1}^{m} k_i\log(p) + (mN - \sum_{i=1}^{m} k_i)\log(1-p)$$

$$\frac{\partial(log-likelihood(p))}{\partial p} = \frac{\sum_{i=1}^{m} k_i}{p} - \frac{mN-\sum_{i=1}^{m} k_i}{1-p} = 0 \;\rightarrow\; (1-p)\sum_{i=1}^{m} k_i - p(mN - \sum_{i=1}^{m} k_i) = \sum_{i=1}^{m} k_i - pmN = 0$$

$$\rightarrow\; p = \frac{\sum_{i=1}^{m} k_i}{mN}$$

Q7:

likelihood=$\frac{1}{2\pi|\Sigma|^{\frac{1}{2}}}\exp\left(\frac{-x_1^t\Sigma^{-1}x_1}{2}\right)\frac{1}{2\pi|\Sigma|^{\frac{1}{2}}}\exp\left(\frac{-x_2^t\Sigma^{-1}x_2}{2}\right)\ldots\frac{1}{2\pi|\Sigma|^{\frac{1}{2}}}\exp(\frac{-x_N^t\Sigma^{-1}x_N}{2})$

Log(likelihood)=$-N\ln(2\pi) - \frac{N}{2}\ln(|\Sigma|) - \frac{1}{2}\sum_{i=1}^{N} x_i^t\Sigma^{-1}x_i$

$$\frac{\partial}{\partial\Sigma}(Log(likelihood)) = -\frac{N}{2}\left(\frac{\frac{\partial|\Sigma|}{\partial\Sigma}}{|\Sigma|}\right) - \frac{1}{2}\sum_{i=1}^{N} x_i^t\frac{\partial\Sigma^{-1}}{\partial\Sigma}x_i = -\frac{N}{2}(\frac{|\Sigma|\Sigma^{-t}}{|\Sigma|}) - \frac{1}{2}\sum_{i=1}^{N}(-\Sigma^{-1})x_ix_i^t\Sigma^{-1}=0$$

$$N = \Sigma^{-1}\sum_{i=1}^{N} x_ix_i^t \;\longrightarrow\; \Sigma = \frac{1}{N}\sum_{i=1}^{N} x_ix_i^t$$

Q8:

likelihood= $\begin{cases} (\frac{1}{1-\theta})^N & \theta \le x_i \le 1 \\ 0 & O.W. \end{cases}$ We want to find a θ that maximizes likelihood. Please notice that if there is a $x_i < \theta$, then the likelihood will be equal to 0. So, we should have $\theta \le \min\{x_i\}$.

On the other hand, we prefer larger value for θ to maximize the above ascending likelihood function. Therefore, $\theta = \min\{x_i\}$ where $i = 1, \dots, N$

Q9:

$$AP = \left(\prod_{i=1}^{N} \frac{\lambda^{x_i}}{x_i!} \exp(-\lambda)\right) \lambda^{\alpha-1} \exp(-\lambda\beta) = \frac{\lambda^{\sum_{i=1}^{N} x_i}}{\prod_{i=1}^{N} x_i!} \exp(-N\lambda)\, \lambda^{\alpha-1} \exp(-\lambda\beta)$$

$$\ln(AP) = \sum_{i=1}^{N} x_i \ln(\lambda) - \ln\left(\prod_{i=1}^{N} x_i!\right) - N\lambda + (\alpha-1)\ln(\lambda) - \lambda\beta$$

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

Q12-

در واقع x[n] خود یک کمیت تصادفی است. میانگین x[n] را محاسبه می کنیم:

$$E\{x[n]\} = E\{As[n]\} + E\{w[n]\} \rightarrow E\{x[n]\} = A\bar{s}$$

در رابطه اخیر فرض بر این است که مقدار $\bar{s}[n]$ معلوم است زیرا s[n] یک سیگنال معلوم است. بنابراین برای تخمین A کافی است $E\{x[n]\}$ را تخمین بزنیم و حاصل را بر $\bar{s}[n]$ تقسیم کنیم. بر اساس رابطه داده شده، $x[n]$ دارای توزیع گوسی با میانگین $A\bar{s}[n]$ و واریانس σ^2 است. بنابراین برای تخمین A به روش بیشینه شباهت کافی است تابع شباهت زیر را تشکیل دهیم:

$$likelihood: \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x[n]-A\bar{s})^2}{2\sigma^2}\right) = 2\pi^{-\frac{N}{2}} \frac{1}{\sigma^{2N}} \exp\left(-\sum_{n=0}^{N-1} \frac{(x[n]-A\bar{s})^2}{2\sigma^2}\right)$$

Log-likelihood: $-\frac{N}{2}\ln(2\pi) - 2N\ln(\sigma) - \sum_{n=0}^{N-1} \frac{(x[n]-A\bar{s})^2}{2\sigma^2}$

MLE: $\frac{\partial \text{Log-likelihood}}{\partial A} = 2\sum_{n=0}^{N-1} \bar{s} \frac{(x[n]-A\bar{s})}{2\sigma^2} = 0 \quad \rightarrow \quad \sum_{n=0}^{N-1} x[n] = A\sum_{n=0}^{N-1} \bar{s} = AN\bar{s} \quad \rightarrow \quad A = \frac{\sum_{n=0}^{N-1} x[n]}{N\bar{s}}$