PML hw3 Theory :

Q1 (a): We have

•) $\log \tilde{p}(x|a,b) = (a-1)\log x - bx$

$\partial_x \log \tilde{p}(x|a,b) = \frac{a-1}{x} - b$ .... setting equal to zero gives mode $\hat{x} = \frac{a-1}{b}$

$$\partial_x^2 \log \tilde{p}(x|a,b)\Big|_{x=\hat{x}} = -\frac{a-1}{\hat{x}^2} = -\frac{a-1}{\left(\frac{a-1}{b}\right)^2} = \frac{-b^2}{a-1} =: \psi$$

The Taylor approximation around $\hat{x}$ is given by:

$$\log \tilde{p}(x|a,b) \simeq \log \tilde{p}(\hat{x}|a,b) - \frac{1}{2}\left(\frac{b^2}{a-1}\right)(x-\hat{x})^2$$

$$\Rightarrow \tilde{p}(x|a,b) = \tilde{p}(\hat{x}|a,b)\exp\left(-\frac{1}{2}\frac{(x-\hat{x})^2}{-\psi^{-1}}\right)$$

We can actually remove the $\sim$ from p since ~~the~~ $\partial_x p = \partial_x \tilde{p}$

So $p(x|a,b) \simeq p(\hat{x}|a,b)\exp\left(-\frac{1}{2}\frac{(x-\hat{x})^2}{-\psi^{-1}}\right)$ $(*)$

which we recognize as the functional form of a normal distribution w/ mean $\frac{a-1}{b}$ and variance $\frac{a-1}{b^2}$

•) Let $b = 1$

from our expression $(*)$, we then have

$$\int_{\mathbb{R}} p(x|a,b) \simeq p(\hat{x}|a,b)\int_{\mathbb{R}}\exp\left(-\frac{1}{2}\frac{(x-\hat{x})^2}{-\psi^{-1}}\right)dx \simeq 1$$ (since p is a pdf)

$$\Rightarrow 1 \simeq \frac{1^a}{\Gamma(a)}\left(\frac{a-1}{1}\right)^{a-1} e^{-(a-1)}\int_{-\infty}^{+\infty}\exp\left(-\frac{(x-(a-1))^2}{2(a-1)}\right)dx$$

$$\Rightarrow \Gamma(a) \simeq (a-1)^{a-1} e^{-(a-1)}\int_{-\infty}^{+\infty}\exp\left(-\frac{(x-(a-1))^2}{2(a-1)}\right)dx$$

$$= (a-1)^{a-1} e^{-(a-1)}\sqrt{2\pi(a-1)}$$ (recognizing the Gaussian Integral)

(b) We consider the product of the Gaussian Likelihood & Wishart prior & show that it also takes the form of a Wishart distribution (unnormalized) ...

We have:

$$\left(\prod_{i=1}^{n} p(\underline{x}_i \mid \Sigma^{-1})\right) p(\Sigma^{-1} \mid W, \vartheta)$$

$$= \left(\frac{1}{(2\pi)^{d/2}} |\Sigma^{-1}|^{1/2}\right)^n \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\underline{x}_i-\underline{\mu})^T\Sigma^{-1}(\underline{x}_i-\underline{\mu})\right) \frac{|\Sigma^{-1}|^{\frac{\vartheta-d-1}{2}}}{2^{\frac{d\vartheta}{2}}|W|^{\vartheta/2}\Gamma_d\left(\frac{\vartheta}{2}\right)} \exp\left(-\frac{1}{2}\mathrm{tr}(W^{-1}\Sigma^{-1})\right)$$

$$\propto |\Sigma^{-1}|^{\frac{\vartheta-d+n-1}{2}} \exp\left(-\frac{1}{2}\left(\left[\sum_{i=1}^{n}(\underline{x}_i-\underline{\mu})^T\Sigma^{-1}(\underline{x}_i-\underline{\mu})\right] + \mathrm{tr}(W^{-1}\Sigma^{-1})\right)\right)$$

We now make use of the following useful trace identity: (since $\mathrm{tr}(AB) = \mathrm{tr}(B^TA^T)$ $\forall A, B \in \mathrm{Mat}_{n\times n}\mathbb{R}$)

- $v^T\Omega v = \mathrm{tr}(\Omega v v^T) = \mathrm{tr}(v v^T \Omega^T)$ $\forall v \in \mathbb{R}^n, \Omega \in \mathrm{Mat}_{n\times n}(\mathbb{R})$

Thus $\forall i \in \{1, \dots, n\}$; we have $(\underline{x}_i-\underline{\mu})^T\Sigma^{-1}(\underline{x}_i-\underline{\mu}) = \mathrm{tr}\left((\underline{x}_i-\underline{\mu})^T(\underline{x}_i-\underline{\mu})\Sigma^{-1}\right)$ (as Σ^{-1} is symmetric)

Since the sum of traces is the trace of a sum we have:

$$\sum_{i=1}^{n}(\underline{x}_i-\underline{\mu})^T\Sigma^{-1}(\underline{x}_i-\underline{\mu}) = \mathrm{tr}\left(\sum_{i=1}^{n}(\underline{x}_i-\underline{\mu})^T(\underline{x}_i-\underline{\mu})\Sigma^{-1}\right)$$

Therefore:

$$\left(\prod p(x_i \mid \Sigma^{-1})\right) p(\Sigma^{-1} \mid W, \vartheta) \propto |\Sigma^{-1}|^{\frac{\vartheta-d+n-1}{2}} \exp\left(-\frac{1}{2}\left[\mathrm{tr}(W^{-1}\Sigma^{-1} + \sum_{i=1}^{n}(\underline{x}_i-\underline{\mu})^T(\underline{x}_i-\underline{\mu})\Sigma^{-1}\right]\right)$$

$$= |\Sigma^{-1}|^{\frac{\vartheta-d+n-1}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\left(W^{-1} + \sum_{i=1}^{n}(\underline{x}_i-\underline{\mu})^T(\underline{x}_i-\underline{\mu})\mathbb{1}\right)\Sigma^{-1}\right)\right]$$

which we recognize as the functional form of the $\mathrm{Wishart}\left(\vartheta+n, \left(W^{-1} + \sum_{i=1}^{n}(\underline{x}_i-\underline{\mu})^T(\underline{x}_i-\underline{\mu})\mathbb{1}\right)^{-1}\right)$ distribution

whence the Wishardt ~~prior~~ is the conjugate prior for Σ^{-1} and the posterior parameters are as above $\blacksquare$