a)

$$p(x|a,b) = g(x;a,b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} \quad \text{with } \Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$$

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

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

$$\frac{\partial \log \tilde{p}}{\partial x} = \frac{a-1}{x} - b = 0 \Rightarrow a-1 = bx \Rightarrow x^* = \frac{a-1}{b} = \hat{x} \text{(mode)}$$

$$\left.\frac{\partial^2 \log \tilde{p}}{\partial x^2}\right|_{x=x^*} = \left.\frac{1-a}{x^2}\right|_{x=x^*} = \frac{(1-a)b^2}{(a-1)^2} = -\frac{b^2}{a-1} \quad \text{and thus } \sigma^2 = \frac{a-1}{b^2}$$

Finally $\underbrace{q(x|a,b)}_{\text{approximation}} = \mathcal{N}\left(x; \frac{a-1}{b}, \frac{a-1}{b^2}\right) = \dfrac{b\, e^{\frac{-b^2\left(x-\frac{a-1}{b}\right)^2}{2(a-1)}}}{\sqrt{2\pi(a-1)}}$

For $b=1$, $q(x|a,b) = \mathcal{N}(x; a-1, a-1) = \dfrac{\exp\left[\frac{(x-a+1)^2}{2(a-1)}\right]}{\sqrt{2\pi(a-1)}}$

To find Stirling approximation:

We know that $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx = (a-1)!$

we also know $\int p(x|a,b) \approx p(\hat{x}) \int \exp\left[\frac{(x-\hat{x})^2}{2\sigma^2}\right] dx = 1$

L03, Slide 35

$$1 \approx \frac{\hat{x}^{a-1} e^{-\hat{x}}}{\Gamma(a)} \int \exp\left[\frac{-(x-\hat{x})^2}{2(a-1)}\right] dx$$

$$1 \approx \frac{(a-1)^{a-1} e^{-(a-1)}}{\Gamma(a)} \sqrt{2\pi(a-1)}$$

$a-1=n$

$$\Gamma(a) \approx (a-1)^{a-1} e^{-(a-1)} \sqrt{2\pi(a-1)} = \boxed{\sqrt{2\pi(a-1)} \left(\frac{a-1}{e}\right)^{a-1}}$$

$\Gamma(n+1) \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$

Q.E.D.

b)

$$\mathcal{N}(x_i;\mu,\Sigma)=\frac{|\Sigma^{-1}|^{\frac{1}{2}}}{(2\pi)^{\frac{d}{2}}}\exp\left[\frac{-(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)}{2}\right]=p(x_i\mid\mu,\Sigma)$$

(or use $p(x_i|\mu,\Sigma^{-1})$)

Show conjugate prior under this likelihood:

$$p(\Sigma^{-1}|W,v)=\mathcal{W}(\Sigma^{-1}|W,v)=\frac{|\Sigma^{-1}|^{(v-d-1)/2}\exp\left(-\frac{tr(W^{-1}\Sigma^{-1})}{2}\right)}{2^{dv/2}|W|^{v/2}\Gamma_d(v/2)}$$

Where tr is the trace and $\Gamma_d=\pi^{d(d-1)/4}\prod_{i=1}^{d}\Gamma\left(z+\frac{1-i}{2}\right)$

We need to prove that the posterior distribution of Σ^{-1} after observing data $\{x_i\}$ is also in the same distributional family as the prior. In other words

$$\underbrace{p(\Sigma^{-1}|W,v,X)}_{\text{Wishart}}\propto p(X|\mu,\Sigma^{-1})\overbrace{p(\Sigma^{-1}|W,v)}^{\text{Wishart}}$$

$$\propto\underbrace{\prod_i p(x_i|\mu,\Sigma^{-1})}_{\text{By i.i.d assumption}}p(\Sigma^{-1}|W,v)$$

It is also a Gaussian

Using the property for positive definite square matrix Σ^{-1}:

$$(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)=tr\left((x_i-\mu)(x_i-\mu)^T\Sigma^{-1}\right)$$

For the multiplication of i.i.d. Gaussians this takes the form:

$$\prod_{i=1}^{n}A_i\exp\left[-tr((x_i-\mu)(x_i-\mu)^T\Sigma^{-1})/2\right]=B\exp\left[-tr\left(\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T\Sigma^{-1}\right)/2\right]$$

Finally multiplying the exponentials of the combined Gaussians and the Wishart prior, we have

$$p(\Sigma^{-1} | W', v', X) \propto \prod_{i=1}^{n \to \text{number of observed points}} p(x_i | \mu, \Sigma^{-1})\, p(\Sigma^{-1} | W, v)$$

$$\propto \frac{|\Sigma^{-1}|^{n}}{(2\pi)^{\frac{nd}{2}}} \exp\left[-tr\left[\left(\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T\right)\Sigma^{-1}/2\right]\right]$$

$$\times \frac{|\Sigma^{-1}|^{(v-d-1)/2}}{2^{vd/2}|W|^{v/2}\Gamma(v/2)} \exp\left[tr\left(W^{-1}\Sigma^{-1}\right)\right]$$

$$\propto \frac{|\Sigma^{-1}|^{(v+n-d-1)/2}}{2^{(v+n)/2}\, C} \exp\left[-tr\left(\left(W^{-1}+\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T\right)\Sigma^{-1}\right)/2\right]$$

where C denotes normalization constant. From this we clearly see that the posterior takes the form:

$$p(\Sigma^{-1} | W', v', X) = p\left(\Sigma^{-1} \middle| \left(W^{-1}+\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T\right)^{-1}, v+n\right)$$

$$= \text{Wishart}\left(\Sigma^{-1} \middle| \left(W^{-1}+\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T\right)^{-1}, v+n\right)$$

Q.E.D

See also: Normal Wishart Distribution