1. 1 The density of $\vec{u}$, is $f_{\vec{u}} = f_{\vec{x}}(g^{-1}(\vec{u}))||J||$

First, we find the partials:

$$\frac{\partial x_i}{\partial u_j} = \frac{1}{u_{d+1}}$$

$$\frac{\partial x_i}{\partial j_i} = \frac{1}{u_i} + \frac{1}{u_{d+1}}$$

Which gives us the Jacobian:

$$J = \begin{bmatrix} \frac{\partial x_1}{\partial u_1} & \cdots & \frac{\partial x_1}{\partial u_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_d}{\partial u_1} & \cdots & \frac{\partial x_n}{\partial u_d} \end{bmatrix} = \begin{bmatrix} \frac{1}{u_1} + \frac{1}{u_{d+1}} & & \frac{1}{u_{d+1}} \\ & \ddots & \\ \frac{1}{u_{d+1}} & & \frac{1}{u_d} + \frac{1}{u_{d+1}} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{u_1} \\ \vdots \\ \frac{1}{u_d} \end{bmatrix} I_d + \begin{bmatrix} \vdots \\ \frac{1}{u_{d+1}} \\ \vdots \end{bmatrix}$$

Letting $X = \begin{bmatrix} \frac{1}{u_1} \\ \vdots \\ \frac{1}{u_d} \end{bmatrix}$ and $C = \begin{bmatrix} \vdots \\ \frac{1}{u_{d+1}} \\ \vdots \end{bmatrix}$, Sylvester's Formula gives us that:

$$\det(J) = \det(X \times I_d) \cdot (1 + CX^{-1} \begin{bmatrix} \ldots & 1 & \ldots \end{bmatrix})$$

$$= \left(\prod_{i=1}^{d} \frac{1}{u_i} \right) \left(1 + \left(\frac{\frac{1}{u_{d+1}}}{\frac{1}{u_1}} + \ldots + \frac{\frac{1}{u_{d+1}}}{\frac{1}{u_d}} \right) \right)$$

$$= \left(\prod_{i=1}^{d} \frac{1}{u_i} \right) \left(\frac{u_1 + \ldots + u_d}{u_{d+1}} + \frac{u_{d+1}}{u_{d+1}} \right)$$

$$= \left(\prod_{i=1}^{d} \frac{1}{u_i} \right) \left(\frac{1 - u_{d+1}}{u_{d+1}} + \frac{u_{d+1}}{u_{d+1}} \right)$$

$$= \left(\prod_{i=1}^{d+1} u_i \right)^{-1}$$

Thus, $||J|| = \left(\prod_{i=1}^{d+1} u_i \right)^{-1}$. We need not worry about the absolute value as each element $u_i \in [0, 1]$.

So we know the density of $\vec{u}$ to be:

$$f_{\vec{u}}(\vec{u}) = |2\pi\Sigma|^{-1/2} \left(\prod_{i=1}^{d+1} u_i\right)^{-1} \exp\left[-\frac{1}{2}\left\{log\left(\frac{u}{u_{d+1}}\right) - \mu\right\}^T \Sigma^{-1} \left\{log\left(\frac{u}{u_{d+1}}\right) - \mu\right\}\right]$$

2. 1 The likelihood function is

$$L(g, \theta_L, \theta_H) = \prod_{i=1}^{N} p(y_{i1}, \ldots, y_{iP}|g, \theta_L, \theta_H)$$

where N and P are defined from the data generating process (number of observations and number of features, respectively). Then,

$$p(y_{i1}, \ldots, y_{iP}|g, \theta_L, \theta_H) = \prod_{j=i}^{P} p(y_{ij}|g, \theta_L, \theta_H)$$

$$p(y_{ij}|g, \theta_L, \theta_H) = g_{H,i} \prod_{k=1}^{d_j} (\theta_{H,jk})^{I(y_{ij}=k)} + g_{L,i} \prod_{k=1}^{d_j} (\theta_{L,jk})^{I(y_{ij}=k)}$$

where d_j is the number of categories of feature j. So we can write the entire likelihood to be:

$$L(g, \theta_L, \theta_H) = \prod_{i=1}^{N}\prod_{j=1}^{P} \left(g_{H,i} \prod_{k=1}^{d_j} (\theta_{H,jk})^{I(y_{ij}=k)} + g_{L,i} \prod_{k=1}^{d_j} (\theta_{L,jk})^{I(y_{ij}=k)}\right)$$

and the log-likelihood:

$$l(g, \theta_L, \theta_H) = \sum_{i=1}^{N}\sum_{j=1}^{P} \left(\log\left(g_{H,i} \prod_{k=1}^{d_j} (\theta_{H,jk})^{I(y_{ij}=k)} + g_{L,i} \prod_{k=1}^{d_j} (\theta_{L,jk})^{I(y_{ij}=k)}\right)\right)$$