PML Theory:

Sam Laing 6283670
Albert Catalan 6443478

a) we have

$$N(x; a, A)\, N(x; b, B)$$

$$= \frac{1}{(2\pi)^{D/2}|A|^{1/2}} \cdot \frac{1}{(2\pi)^{D/2}|B|^{1/2}} \exp\left[-\frac{1}{2}\left[(x-a)^T A^{-1}(x-a) + (x-b)^T B^{-1}(x-b)\right]\right]$$

$$\propto \exp\left[-\frac{1}{2}\left[x^T(A^{-1}+B^{-1})x - 2(A^{-1}a+B^{-1}b)^T x + a^T A^{-1} a + b^T B^{-1} b\right]\right]$$

completing the square

$$= \exp\left[-\frac{1}{2}\left[(x-(A^{-1}+B^{-1})^{-1}(A^{-1}a+B^{-1}b))^T(A^{-1}+B^{-1})(x-(A^{-1}+B^{-1})^{-1}(A^{-1}a+B^{-1}b)) - (A^{-1}a+B^{-1}b)^T(A^{-1}+B^{-1})^{-1}(A^{-1}a+B^{-1}b) + a^T A^{-1} a + b^T B^{-1} b\right]\right]$$

$$\propto \mathcal{N}(x; \overbrace{(A^{-1}+B^{-1})^{-1}(A^{-1}a+B^{-1}b)}^{\mu_*}, \overbrace{(A^{-1}+B^{-1})^{-1}}^{\Sigma_*}) \cdot \exp\left[-\frac{1}{2}\left[a^T A^{-1} a + b^T B^{-1} b - (a^T A^{-1} + b^T B^{-1})(A^{-1}+B^{-1})^{-1}(A^{-1}a+B^{-1}b)\right]\right]$$

$$= \mathcal{N}(x; \mu_*, \Sigma_*) \exp\left(-\frac{1}{2}\left[a^T A^{-1} a + b^T B^{-1} b \cancel{+(a^T A^T + b^T B^{-1})} - a^T A^{-1}(A^{-1}+B^{-1})^{-1}A^{-1}a - b^T B^{-1}(A^{-1}+B^{-1})^{-1}B^{-1}b - 2b^T B^{-1}(A^{-1}+B^{-1})^{-1}A^{-1}a\right]\right)$$

$$\propto \mathcal{N}(x; \mu_*, \Sigma_*) \exp\left(-\frac{1}{2}\left[a^T(A^{-1} - A^{-1}(A^{-1}+B^{-1})^{-1}A^{-1})a - b^T(B^{-1} - B^{-1}(A^{-1}+B^{-1})B^{-1})b\right]\right)$$

(where other terms are absorbed by proportionality)

$$\propto \mathcal{N}(x; \mu_*, \Sigma_*) \exp\left[-\frac{1}{2}(a-b)^T(A+B)^{-1}(a-b)\right]$$

woodbury identity

$$\propto N(x; (A^{-1}+B^{-1})^{-1}(A^{-1}a+B^{-1}b), (A^{-1}+B^{-1})^{-1}) \cdot N(a, b, A+B) \quad \blacksquare$$

b) for a matrix $\Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}$, ~~where~~ we have

$$\Sigma^{-1} = \Sigma^{*} = \begin{pmatrix} \Sigma^{*}_{xx} & \Sigma^{*}_{xy} \\ \Sigma^{*}_{yx} & \Sigma^{*}_{yy} \end{pmatrix}$$

where $\Sigma^{*}_{xx} = A^{-1}$ $\quad \Sigma^{*}_{xy} = -A^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}$

$\Sigma^{*}_{yx} = -\Sigma_{yy}^{-1}\Sigma_{yx}A^{-1}$ $\quad \Sigma_{yy}^{*} = \Sigma_{yy}^{-1} + \Sigma_{yy}^{-1}\Sigma_{yx}A^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}$

with $A := \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}$

(by Schur's identity)

Now let $z^T := (x^T, y^T)$ & $\mu^T = (\mu_x^T, \mu_y^T)$

we have:

$$(z-\mu)^T\Sigma^{-1}(z-\mu) = (x^T-\mu_x^T, y^T-\mu_y^T)\begin{pmatrix} \Sigma^{*}_{xx} & \Sigma^{*}_{xy} \\ \Sigma^{*}_{yx} & \Sigma^{*}_{yy} \end{pmatrix}\begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix}$$

$$= (x-\mu_x)^T\Sigma^{*}_{xx}(x-\mu_x) + (x-\mu_x)^T\Sigma^{*}_{xy}(y-\mu_y)$$

$$+ (y-\mu_y)^T\Sigma^{*}_{yx}(x-\mu_x) + (y-\mu_y)^T\Sigma^{*}_{yy}(y-\mu_y)$$

$$= (x-\mu_x)^T A^{-1}(x-\mu_x) - (x-\mu_x)^T(A^{-1}\Sigma_{xy}\Sigma_{yy}^{-1})(y-\mu_y)$$

$$- (y-\mu_y)^T\Sigma_{yy}^{-1}\Sigma_{yx}A^{-1}(x-\mu_x)$$

$$+ (y-\mu_y)^T\left[\Sigma_{yy}^{-1} + \Sigma_{yy}^{-1}\Sigma_{yx}A^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}\right](y-\mu_y)$$

$$= (y-\mu_y)^T \Sigma_{yy}^{-1} (y-\mu_y)$$

$$+ (x-(\mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y-\mu_y)))^T A^{-1} (x-(\mu_x+\Sigma_{xy}\Sigma_{yy}^{-1}(y-\mu_y)))$$

now we have

$$p(x|y) = \frac{p(x,y)}{p(y)} \qquad // \; z^T = (x^T, y^T)$$

$$\propto \exp\left(-\frac{1}{2}\left[(z-\mu)^T \Sigma^{-1} (z-\mu) - (y-\mu_y)^T \Sigma_{yy}^{-1}(y-\mu_y)\right]\right)$$

$$= \exp\left[-\frac{1}{2}\left[(x-(\mu_x+\Sigma_{xy}\Sigma_{yy}^{-1}(y-\mu_y)))^T A^{-1} (x-(\mu_x+\Sigma_{xy}\Sigma_{yy}^{-1}(y-\mu_y)))\right]\right]$$

using derived expression

which we recognize as the functional form of a $\mathcal{N}(\mu_x+\Sigma_{xy}\Sigma_{yy}^{-1}(y-\mu_y), \Sigma_{xx}-\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})$

~~c) Consider a multivariate normal $\mathcal{N}(z; \mu_z, \Sigma_z)$ where~~

~~$\mu_z^T = (\mu_1, \mu_2, \mu^T)$ $\quad \Sigma_z = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{XY} \\ \Sigma_{21} & \Sigma_{22} & \\ \Sigma_{YX} & & \Sigma_{YY} \end{pmatrix}$~~

~~where $\mu \in \mathbb{R}^{n-2}$, $\Sigma_{YY} \in \mathbb{R}^{(n-2)\times(n-2)}$~~

~~$\mu_z \in \mathbb{R}^n$, $\Sigma_z \in \mathbb{R}^{n\times n}$ (since wlog, $(i,j)=(1,2)$)~~

c) •) WLOG suppose $z = \begin{pmatrix} x_1 \\ x_2 \\ y \end{pmatrix}$ $\Sigma = \left(\begin{array}{cc|c} \sigma_{11} & \sigma_{12} & \Sigma_{xy} \\ \sigma_{21} & \sigma_{22} & \\ \hline \Sigma_{yx} & & \Sigma_{yy} \end{array}\right)$

$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_y \end{pmatrix}$

then $p(x_1, x_2) = \mathcal{N}\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}; \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}\right)$ (*)

We have (using the result from (b))

$$p(x_1|x_2) = \mathcal{N}\left(x_1; \mu_1 + \sigma_{12} \frac{1}{\sigma_{22}}(x_2 - \mu_2), \sigma_{11} - \sigma_{12} \frac{1}{\sigma_{22}} \sigma_{21}\right)$$

$$= \mathcal{N}(x; \mu_1, \sigma_{11}) = p(x_1) \quad \blacksquare$$

(if $\sigma_{12} = 0$)

(*) The reason for this is that

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 \cdots & 0 \end{pmatrix}}_{A} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$\cancel{\Rightarrow p(x_1, x_2) = A\mu = \mu_1}$

$\Rightarrow A\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ $\cancel{\& A \Sigma A^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \cdots 0 \\ 0 & 1 & 0 & \cdots 0 \end{pmatrix} \begin{pmatrix} \sigma_{11} & \\ & \sigma_{nn} \end{pmatrix}}$

$\& A \Sigma A^{-1} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$ (with A clearly linear)

$\Rightarrow (X_1, X_2) \sim \mathcal{N}\left((\mu_1, \mu_2), \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}\right)$

(linear / affine transformations of normal are normal

•) As in part 1, we can linearly project to see that

$$p(x_1, x_2, x_3) = \mathcal{N}\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \underbrace{\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}}_{\mu}, \underbrace{\begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}}_{\Sigma} \right)$$

(where wlog $x_k = x_3$) (Using matrix $A = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}$)

Since Σ is symmetric, so is Σ^{-1}.
The ij^{th} entry of Σ^{-1} is the ij^{th} cofactor divided by the determinant

therefore, if $(\Sigma^{-1})_{12} = 0$, $(\Sigma^{-1})_{21} = 0$

then $\begin{vmatrix} \sigma_{21} & \sigma_{23} \\ \sigma_{31} & \sigma_{33} \end{vmatrix} = \sigma_{21}\sigma_{33} - \sigma_{31}\sigma_{23} = 0$

$$\Rightarrow \sigma_{21} = \frac{\sigma_{31}\sigma_{23}}{\sigma_{33}} \quad (*)$$

and $\begin{vmatrix} \sigma_{12} & \sigma_{13} \\ \sigma_{31} & \sigma_{33} \end{vmatrix} = \sigma_{12}\sigma_{33} - \sigma_{13}\sigma_{31} = 0$

$$\Rightarrow \sigma_{12} = \frac{\sigma_{13}\sigma_{31}}{\sigma_{33}} \quad (*)$$

Then, using part (b),

$$p(x_1, x_2 \mid x_3) = \mathcal{N}\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}; \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \sigma_{13} \\ \sigma_{23} \end{pmatrix} \frac{1}{\sigma_{33}}(x_3 - \mu_3), \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} - \frac{1}{\sigma_{33}} \begin{pmatrix} \sigma_{13}\sigma_{31} & \sigma_{13}\sigma_{31} \\ \sigma_{23}\sigma_{31} & \sigma_{23}\sigma_{32} \end{pmatrix} \right)$$

$$\cancel{= \mathcal{N}\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \begin{pmatrix} \mu_1 + \frac{\sigma_{13}}{\sigma_{33}}(x_3 - \mu_3) \\ \mu_2 + \frac{\sigma_{23}}{\sigma_{33}}(x_3 - \mu_3) \end{pmatrix} \left(\sigma_{11} - \frac{\sigma_{13}\sigma_{31}}{} \right. \right.}$$

$$= N\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}; \begin{pmatrix} \mu_1 + \frac{\sigma_{13}}{\sigma_{33}}(x_3-\mu_3) \\ \mu_2 + \frac{\sigma_{23}}{\sigma_{33}}(x_3-\mu_3) \end{pmatrix}, \begin{pmatrix} \sigma_{11} - \frac{\sigma_{13}\sigma_{31}}{\sigma_{33}} & \sigma_{12} - \frac{\sigma_{13}\sigma_{32}}{\sigma_{33}} \\ \sigma_{21} - \frac{\sigma_{23}\sigma_{31}}{\sigma_{33}} & \sigma_{22} - \frac{\sigma_{23}\sigma_{32}}{\sigma_{33}} \end{pmatrix}\right)$$

but using our identities $(*)$,
we see that the diagonals of the covariance matrix must then be 0

$$\Rightarrow p(x_1, x_2 | x_3) = N\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}; \begin{pmatrix} \mu_1 + \frac{\sigma_{13}}{\sigma_{33}}(x_3-\mu_3) \\ \mu_2 + \frac{\sigma_{23}}{\sigma_{33}}(x_3-\mu_3) \end{pmatrix}, \begin{pmatrix} \sigma_{11} - \frac{\sigma_{13}\sigma_{31}}{\sigma_{33}} & 0 \\ 0 & \sigma_{11} - \frac{\sigma_{23}\sigma_{32}}{\sigma_{33}} \end{pmatrix}\right)$$

but then we see that (reading the matrix)

$$p(x_1 | x_3) = N\left(x_1 ; \mu_1 + \frac{\sigma_{13}}{\sigma_{33}}(x_3-\mu_3), \sigma_{11} - \frac{\sigma_{13}\sigma_{31}}{\sigma_{33}}\right)$$

$$p(x_2 | x_3) = N\left(x_2 ; \mu_2 + \frac{\sigma_{23}}{\sigma_{33}}(x_3-\mu_3), \sigma_{22} - \frac{\sigma_{23}\sigma_{32}}{\sigma_{33}}\right)$$

which are precisely the formulae ~~of~~ ~~for~~ from part (b) when applied to $\begin{pmatrix} x_1 \\ x_3 \end{pmatrix}$ & $\begin{pmatrix} x_2 \\ x_3 \end{pmatrix}$

We therefore conclude that x_1 & x_2 are conditionally independent given x_k for any $k \in \{3, \dots, n\}$ $\blacksquare$