MATH 559 - EXERCISES 2 : SOLUTIONS

1. *Suppose $Y_1, \ldots, Y_n$ are realizations from an exchangeable binary sequence. Using the Jeffreys prior for parameter*

$$\theta = \mathbb{E}_Y[Y]$$

find an approximation to the posterior distribution $\pi_n(\theta)$ for large n.

We have

$$\ell_n(\theta) = \sum_{i=1}^{n} \log f_Y(y_i;\theta) = \sum_{i=1}^{n}((1-y_i)\log(1-\theta) + y_i \log\theta)$$

$$\dot{\ell}_n(\theta) = -\frac{n-s_n}{1-\theta} + \frac{s_n}{\theta} \qquad s_n = \sum_{i=1}^{n} y_i$$

$$\ddot{\ell}_n(\theta) = -\left[\frac{(n-s_n)}{(1-\theta)^2} + \frac{s_n}{\theta^2}\right]$$

which informs us that as $\mathbb{E}[S_n] = n\theta$, the Jeffreys prior is

$$\pi_0(\theta) \propto \left|\mathbb{E}\left[-\ddot{\ell}_n(\theta)\right]\right|^{1/2} \propto \frac{1}{\sqrt{\theta(1-\theta)}} \equiv Beta(1/2, 1/2)$$

but also that, for large n, using the results concerning approximations of the likelihood

$$\pi_n(\theta) \ \propto \ \exp\left\{-\frac{n}{2\widehat{\theta}_n(1-\widehat{\theta}_n)}(\theta - \widehat{\theta}_n)^2\right\}$$

as we can **ignore the influence of the prior** when n is large. Thus in this case

$$\pi_n(\theta) \approx Normal(\widehat{\theta}_n, \widehat{\theta}_n(1-\widehat{\theta}_n)/n)$$

2. *Suppose that in a Bayesian model, we have that*

$$f_Y(y;\theta) = \mathbb{1}_{(\theta,\infty)}(y)\exp\{-(y-\theta)\} \quad y \in \mathbb{R}$$

for $\theta \in \Theta \equiv \mathbb{R}^+$. Suppose that the prior is $\pi_0(\theta) \equiv Exponential(2)$. Find the posterior, $\pi_n(\theta)$, based on a sample $y_1, \ldots, y_n$.

We have for the likelihood

$$\mathcal{L}_n(\theta) = \prod_{i=1}^{n} \mathbb{1}_{(\theta,\infty)}(y_i)\exp\{-(y_i-\theta)\} = \mathbb{1}_{(0,\min_i y_i)}(\theta)\exp\left\{-\sum_{i=1}^{n}(y_i-\theta)\right\}$$

$$\propto \ \exp\{n\theta\} \qquad 0 < \theta < \min_i y_i$$

so if $y_{\min}$ is the minimum observed y value, for the posterior, we have

$$\pi_n(\theta) \ \propto \ \exp\{n\theta\}\exp\{-2\theta\} = \exp\{(n-2)\theta\} \qquad 0 < \theta < y_{\min}$$

The normalizing constant is easily computed by integration: If $n > 2$

$$\int_0^{y_{\min}} \exp\{(n-2)\theta\}\, d\theta = \frac{1}{(n-2)}\left[\exp\{(n-2)y_{\min}\} - 1\right]$$

and hence

$$\pi_n(\theta) = \frac{(n-2)}{\exp\{(n-2)y_{\min}\} - 1} \exp\{(n-2)\theta\} \qquad 0 < \theta < y_{\min}$$

and zero otherwise. If $n = 2$, we have

$$\pi_n(\theta) \propto 1 \qquad 0 < \theta < y_{\min}$$

so the posterior distribution is $Uniform(0, y_{\min})$. If $n = 1$,

$$\pi_n(\theta) = \frac{1}{1 - \exp\{-y_1\}} \exp\{-\theta\} \qquad 0 < \theta < y_1.$$

3. *Suppose that in a Bayesian model, we have that*

$$f_Y(y;\theta) = \mathbb{1}_{(0,\infty)}(y) \frac{y}{\theta^2} \exp\left\{-\frac{y^2}{2\theta^2}\right\} \quad y \in \mathbb{R}$$

for $\theta \in \Theta \equiv \mathbb{R}^+$. Using a prior of your choosing, find the posterior, $\pi_n(\theta)$ based on a sample $y_1, \ldots, y_n$.

We have for the likelihood

$$\begin{aligned}\mathcal{L}_n(\theta) &= \prod_{i=1}^n \mathbb{1}_{(0,\infty)}(y_i) \frac{y_i}{\theta^2} \exp\left\{-\frac{y_i^2}{2\theta^2}\right\} \\ &\propto \left(\frac{1}{\theta^2}\right)^n \exp\left\{-\frac{t}{2\theta^2}\right\} \qquad \theta > 0 \qquad\qquad t = \sum_{i=1}^n y_i^2\end{aligned}$$

Writing $\phi = \theta^2$, a conjugate prior for this likelihood is

$$\frac{1}{\phi} \sim Gamma(a_0, b_0/2)$$

that is $\phi \sim InvGamma(a_0, b_0/2)$, with

$$\pi_0(\phi) = \frac{b_0^{a_0}}{\Gamma(a_0)} \left(\frac{1}{\phi}\right)^{a_0+1} \exp\left\{-\frac{b_0}{2\phi}\right\}.$$

Then we have

$$\pi_n(\phi) \equiv InvGamma(n + a_0, (t + b_0)/2)$$

from which we may deduce the posterior for θ by transformation.

4. *Suppose exchangeable sequences $\{Y_{1n}, Y_{2n}\}$ are such that given parameters $\theta_1, \theta_2, \sigma^2$*

$$Y_{ji} \sim Normal(\theta_j, \sigma^2) \qquad j = 1, 2,\ i = 1, \ldots, n_j$$

are independent. Suppose that a proper, conjugate prior specification with

$$\pi_0(\theta_1, \theta_2, \sigma^2) = \pi_0(\sigma^2)\pi_0(\theta_1|\sigma^2)\pi_0(\theta_2|\sigma^2)$$

is used. Compute the posterior distribution for

$$\psi = \theta_2 - \theta_1.$$

From lectures, and by conditional independence, we know that, conditional on σ^2,

$$\pi_n(\theta_1, \theta_2|\sigma^2) = \pi_{n_1}(\theta_1|\sigma^2)\pi_{n_2}(\theta_2|\sigma^2)$$

with

$$\pi_{n_j}(\theta_j|\sigma^2) \equiv Normal\left(\eta_{n_j}, \sigma^2/\lambda_{n_j}\right) \qquad j = 1, 2.$$

where, for $j = 1, 2$,

$$\eta_{n_j} = \frac{n_j\overline{y}_{jn_j} + \lambda_j\eta}{n_j + \lambda_j} \qquad\qquad \lambda_{n_j} = n_j + \lambda_j.$$

Therefore, by properties of the Normal distribution

$$\pi_n(\psi|\sigma^2) \equiv Normal(\eta_n, \sigma^2/\lambda_n)$$

where

$$\eta_n = \eta_{n_2} - \eta_{n_1} \qquad \lambda_n = \frac{\lambda_{n_1}\lambda_{n_2}}{\lambda_{n_1} + \lambda_{n_2}}.$$

Therefore, under the conjugate $InvGamma(a_0/2, b_0/2)$ prior for σ^2, we have from lectures that $\pi_n(\psi)$ is a Student-t distribution

$$\pi_n(\psi) = \frac{\Gamma((a_n+1)/2)}{\Gamma(a_n/2)\sqrt{\pi}} \frac{1}{a_n^{1/2}} \left(\frac{1}{\phi_n}\right)^{1/2} \left\{1 + \frac{1}{a_n}\frac{(\psi - \eta_n)^2}{\phi_n}\right\}^{-(a_n+1)/2}$$

where

$$a_n = n_1 + n_2 + a_0$$

$$b_n = \frac{n_1\lambda_1}{n_1 + \lambda_1}(\overline{y}_{n_1} - \eta_1)^2 + \sum_{i=1}^{n_1}(y_{1i} - \overline{y}_{1n})^2 + \frac{n_2\lambda_2}{n_2 + \lambda_2}(\overline{y}_{2n_2} - \eta_2)^2 + \sum_{i=1}^{n_2}(y_{2i} - \overline{y}_{2n_2})^2 + b_0$$

and where

$$\phi_n = \frac{b_n}{a_n\lambda_n}$$

5. *Suppose exchangeable sequences $\{\mathbf{Y}_n\}$ are assumed to arise from a Bayesian model with*

$$f_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\theta}) \equiv Normal_2(\boldsymbol{\theta}, \Sigma_0)$$

where $\mathbf{Y}_1, \ldots \mathbf{Y}_n$ are 2×1 random vectors that are conditionally independent given parameters $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top$, where Σ_0 is a known covariance matrix.

(i) *Find the posterior distribution for $\boldsymbol{\theta}$ if a conjugate prior is used.*
(ii) *Find the marginal posteriors for θ_1 and for θ_2.*
(iii) *Find the conditional posterior for θ_2 given θ_1.*

(i) Up to proportionality, the likelihood in this case, using the bivariate Normal distribution pdf, takes the form

$$\mathcal{L}_n(\boldsymbol{\theta}) \propto \prod_{i=1}^n \exp\left\{-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\theta})^\top\Sigma_0^{-1}(\mathbf{y}_i - \boldsymbol{\theta})\right\}.$$

The term in the exponent resulting from the product can be written

$$\sum_{i=1}^n(\mathbf{y}_i - \boldsymbol{\theta})^\top\Sigma_0^{-1}(\mathbf{y}_i - \boldsymbol{\theta}) = \sum_{i=1}^n(\mathbf{y}_i - \overline{\mathbf{y}}_n)^\top\Sigma_0^{-1}(\mathbf{y}_i - \overline{\mathbf{y}}_n) + n(\boldsymbol{\theta} - \overline{\mathbf{y}}_n)^\top\Sigma_0^{-1}(\boldsymbol{\theta} - \overline{\mathbf{y}}_n)$$

using the usual sum-of-squares decomposition. A conjugate prior is therefore the $Normal_2(\mathbf{m}_0, \mathbf{M}_0)$

$$\pi_0(\boldsymbol{\theta}) \ \propto \ \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}-\mathbf{m}_0)^\top \mathbf{M}_0^{-1}(\boldsymbol{\theta}-\mathbf{m}_0)\right\}.$$

To compute the posterior, we first note that in the exponent, combining two terms using the complete-the-square formula, we have

$$n(\boldsymbol{\theta}-\overline{\mathbf{y}}_n)^\top \Sigma_0^{-1}(\boldsymbol{\theta}-\overline{\mathbf{y}}_n) + (\boldsymbol{\theta}-\mathbf{m}_0)^\top \mathbf{M}_0^{-1}(\boldsymbol{\theta}-\mathbf{m}_0) = (\boldsymbol{\theta}-\mathbf{m}_n)^\top \mathbf{M}_n^{-1}(\boldsymbol{\theta}-\mathbf{m}_n) + c_n$$

where

$$\mathbf{M}_n = \left(n\Sigma_0^{-1} + \mathbf{M}_0^{-1}\right)^{-1} \qquad \mathbf{m}_n = \left(n\Sigma_0^{-1} + \mathbf{M}_0^{-1}\right)^{-1}\left(n\Sigma_0^{-1}\overline{\mathbf{y}}_n + \Sigma_0^{-1}\mathbf{m}_0\right).$$

Thus we conclude that

$$\pi_n(\boldsymbol{\theta}) \ \propto \ \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}-\mathbf{m}_n)^\top \mathbf{M}_n^{-1}(\boldsymbol{\theta}-\mathbf{m}_n)\right\}$$

and so $\pi_n(\boldsymbol{\theta}) \equiv Normal_2(\mathbf{m}_n, \mathbf{M}_n)$.

(ii) By properties of the multivariate Normal distribution (see Appendix), we have that if

$$\mathbf{m}_n = \begin{pmatrix} m_{n1} \\ m_{n2} \end{pmatrix} \qquad \mathbf{M}_n = \begin{pmatrix} M_{n11} & M_{n12} \\ M_{n21} & M_{n22} \end{pmatrix}$$

then

$$\pi_n(\theta_1) \equiv Normal(m_{n1}, M_{n11}) \qquad \pi_n(\theta_2) \equiv Normal(m_{n2}, M_{n22})$$

(iii) By properties of the multivariate Normal distribution, we have that

$$\pi_n(\theta_2|\theta_1) \equiv Normal(m_{n2} + (M_{n21}(\theta_1 - m_{n1})/M_{n11}), M_{n22} - M_{n12}^2/M_{n11})$$

6. *Show that, in general, Bayes estimators defined by expected loss minimization are not invariant to 1-1 transformations; that is, if $\widehat{\theta}_{nB}$ is a Bayes estimator of θ, and $\phi = g(\theta)$ is 1-1 reparameterization of the model, then*

$$\widehat{\phi}_{nB} \neq g(\widehat{\theta}_{nB})$$

in general.

A counterexample suffices to demonstrate that the result does not hold in general. Suppose that $\theta > 0$. We have that

$$\widehat{\theta}_{nB} = \arg\min_t \int L_\theta(t,\theta)\pi_n(\theta)\, d\theta.$$

and under quadratic loss, $L_\theta(t,\theta) = (t-\theta)^2$, we have seen that the estimate is the posterior mean

$$\widehat{\theta}_{nB} = \mathbb{E}_{\pi_n}[\theta].$$

Now suppose $\phi = g(\theta) = \theta^2$, so that $g(x) = x^2$, and $\theta = \sqrt{\phi}$. We must specify the loss to be the same for a given ϕ as it would be for the corresponding θ, that is

$$L_\phi(t,\phi) \equiv L_\theta(t,\theta) \qquad \theta = \sqrt{\phi}.$$

Hence we must have

$$L_\phi(t,\phi) = (t - \sqrt{\phi})^2$$

We conclude by the usual method that $\widehat{\phi}_{nB} = \mathbb{E}_{\pi_n^\phi}[\sqrt{\phi}]$ computed under the posterior for ϕ. But in general

$$\mathbb{E}_{\pi_n^\phi}[\sqrt{\phi}] \equiv \mathbb{E}_{\pi_n}[\theta] \neq \{\mathbb{E}_{\pi_n}[\theta]\}^2$$

by standard arguments.

APPENDIX

CALCULATIONS FOR THE MULTIVARIATE NORMAL DISTRIBUTION

The **multivariate Normal distribution** is a multivariate generalization of the Normal distribution. The joint pdf of $\mathbf{X} = (X_1, \ldots, X_d)^\top$ takes the form

$$f_{X_1,\ldots,X_d}(x_1, \ldots, x_d) = \left(\frac{1}{2\pi}\right)^{d/2} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

where $\mathbf{x} = (x_1, \ldots, x_d)^\top$, $\boldsymbol{\mu}$ is a $d \times 1$ vector, and Σ is a symmetric, positive-definite $d \times d$ matrix. The distribution is obtained by taking a vector $\mathbf{Z} = (Z_1, \ldots, Z_d)^\top$ of independent standard Normal random variables with joint pdf

$$f_{Z_1,\ldots,Z_d}(z_1, \ldots, z_d) = \left(\frac{1}{2\pi}\right)^{d/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{d} z_i^2\right\} = \left(\frac{1}{2\pi}\right)^{d/2} \exp\left\{-\frac{1}{2}\mathbf{z}^\top\mathbf{z}\right\}$$

and taking the linear transformation

$$\mathbf{X} = \mathbf{L}\mathbf{Z} + \boldsymbol{\mu}$$

where $\mathbf{L}$ is the Cholesky factor of Σ, that is,

$$\Sigma = \mathbf{L}\mathbf{L}^\top.$$

Using the multivariate transformation result, we can deduce the multivariate Normal joint pdf. It can be shown that for any linear combination

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$$

for constant matrix $\mathbf{A}$ and vector $\mathbf{b}$ (compatible in dimension) also has a multivariate Normal distribution; this result can be derived using moment generating functions; we have for $\mathbf{t} = (t_1, \ldots, t_d)^\top \in \mathbb{R}^d$, by independence

$$M_{\mathbf{Z}}(\mathbf{t}) = \exp\left\{\frac{1}{2}\sum_{i=1}^{d} t_i^2\right\} = \exp\left\{\frac{1}{2}\mathbf{t}^\top\mathbf{t}\right\}$$

so therefore

$$\begin{aligned}
M_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}_{\mathbf{X}}[\exp\{\mathbf{t}^\top\mathbf{X}\}] &= \mathbb{E}_{\mathbf{Z}}[\exp\{\mathbf{t}^\top(\mathbf{L}\mathbf{Z}+\boldsymbol{\mu})\}] \\
&= \exp\{\mathbf{t}^\top\boldsymbol{\mu}\}\mathbb{E}_{\mathbf{Z}}[\exp\{(\mathbf{t}^\top\mathbf{L})\mathbf{Z})\}] \\
&= \exp\{\mathbf{t}^\top\boldsymbol{\mu}\}M_{\mathbf{Z}}(\mathbf{L}^\top\mathbf{t}) \\
&= \exp\{\mathbf{t}^\top\boldsymbol{\mu}\}\exp\left\{\frac{1}{2}(\mathbf{L}^\top\mathbf{t})^\top(\mathbf{L}^\top\mathbf{t})\right\} \\
&= \exp\left\{\mathbf{t}^\top\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^\top(\mathbf{L}\mathbf{L}^\top)\mathbf{t}\right\} \\
&= \exp\left\{\mathbf{t}^\top\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^\top\Sigma\mathbf{t}\right\}.
\end{aligned}$$

The distribution of $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ can be deduced using similar methods as

$$\mathbf{Y} \sim Normal_d(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\Sigma\mathbf{A}^\top).$$

Marginal And Conditional Distributions

All marginal and all conditional distributions derived from the multivariate Normal are also multivariate normal; for the marginal distributions, the result follows immediately from the derivation above Suppose that vector random variable $\mathbf{X} = (X_1, X_2, \ldots, X_d)^\top$ has a multivariate normal distribution with pdf given by

$$f_{\mathbf{X}}(\mathbf{x}) = \left(\frac{1}{2\pi}\right)^{d/2} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{x}^\top \Sigma^{-1} \mathbf{x}\right\} \tag{1}$$

where Σ is the $d \times d$ variance-covariance matrix (we can consider here the case where the expected value μ is the $d \times 1$ zero vector; results for the general case are easily available by transformation).

Consider partitioning $\mathbf{X}$ into two components $\mathbf{X}_1$ and $\mathbf{X}_2$ of dimensions d_1 and $d_2 = d - d_1$ respectively, that is,

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}.$$

We attempt to deduce

(a) the **marginal** distribution of $\mathbf{X}_1$, and

(b) the **conditional** distribution of $\mathbf{X}_2$ **given** that $\mathbf{X}_1 = \mathbf{x}_1$.

First, write

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

where Σ_{11} is $d_1 \times d_1$, Σ_{22} is $d_2 \times d_2$, $\Sigma_{21} = \Sigma_{12}^\top$, and

$$\Sigma^{-1} = \mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix}$$

so that $\Sigma\mathbf{V} = \mathbf{I}_d$ ($\mathbf{I}_r$ is the $r \times r$ identity matrix) gives

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{d_1} & 0 \\ 0 & \mathbf{I}_{d_2} \end{bmatrix}$$

where 0 represents the zero matrix of appropriate dimension. More specifically,

$$\begin{aligned} \Sigma_{11}\mathbf{V}_{11} + \Sigma_{12}\mathbf{V}_{21} &= \mathbf{I}_{d_1} & (2)\\ \Sigma_{11}\mathbf{V}_{12} + \Sigma_{12}\mathbf{V}_{22} &= 0 & (3)\\ \Sigma_{21}\mathbf{V}_{11} + \Sigma_{22}\mathbf{V}_{21} &= 0 & (4)\\ \Sigma_{21}\mathbf{V}_{12} + \Sigma_{22}\mathbf{V}_{22} &= \mathbf{I}_{d_2}. & (5) \end{aligned}$$

From the multivariate normal pdf in equation (1), we can re-express the term in the exponent as

$$\mathbf{x}^\top \Sigma^{-1} \mathbf{x} = \mathbf{x}_1^\top \mathbf{V}_{11}\mathbf{x}_1 + \mathbf{x}_1^\top \mathbf{V}_{12}\mathbf{x}_2 + \mathbf{x}_2^\top \mathbf{V}_{21}\mathbf{x}_1 + \mathbf{x}_2^\top \mathbf{V}_{22}\mathbf{x}_2. \tag{6}$$

In order to compute the marginal and conditional distributions, we must complete the square in $\mathbf{x}_2$ in this expression. We can write

$$\mathbf{x}^\top \Sigma^{-1} \mathbf{x} = (\mathbf{x}_2 - \mathbf{m})^\top \mathbf{M}(\mathbf{x}_2 - \mathbf{m}) + \mathbf{c} \tag{7}$$

and by comparing with equation (6) we can deduce that, for quadratic terms in $\mathbf{x}_2$,

$$\mathbf{x}_2^\top \mathbf{V}_{22}\mathbf{x}_2 = \mathbf{x}_2^\top \mathbf{M}\mathbf{x}_2 \qquad \therefore \qquad \mathbf{M} = \mathbf{V}_{22}$$

for linear terms

$$\mathbf{x}_2^\top \mathbf{V}_{21}\mathbf{x}_1 = -\mathbf{x}_2^\top \mathbf{M}\mathbf{m} \qquad \therefore \qquad \mathbf{m} = -\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{x}_1$$

and for constant terms

$$\mathbf{x}_1^\top \mathbf{V}_{11}\mathbf{x}_1 = \mathbf{c} + \mathbf{m}^\top \mathbf{M}\mathbf{m} \qquad \therefore \qquad \mathbf{c} = \mathbf{x}_1^\top(\mathbf{V}_{11} - \mathbf{V}_{21}^\top \mathbf{V}_{22}^{-1}\mathbf{V}_{21})\mathbf{x}_1$$

thus yielding all the terms required for equation (7), that is

$$\mathbf{x}^\top \Sigma^{-1}\mathbf{x} = (\mathbf{x}_2 + \mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{x}_1)^\top \mathbf{V}_{22}(\mathbf{x}_2 + \mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{x}_1) + \mathbf{x}_1^\top(\mathbf{V}_{11} - \mathbf{V}_{21}^\top \mathbf{V}_{22}^{-1}\mathbf{V}_{21})\mathbf{x}_1,$$

which, crucially, is a sum of two terms, where the first can be interpreted as a function of $\mathbf{x}_2$, given $\mathbf{x}_1$, and the second is a function of $\mathbf{x}_1$ only.

Hence we have a factorization of the joint pdf using the chain rule for random variables;

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2|\mathbf{x}_1)f_{\mathbf{X}_1}(\mathbf{x}_1) \tag{8}$$

where

$$f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2|\mathbf{x}_1) \propto \exp\left\{-\frac{1}{2}(\mathbf{x}_2 + \mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{x}_1)^\top \mathbf{V}_{22}(\mathbf{x}_2 + \mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{x}_1)\right\}$$

giving that

$$\mathbf{X}_2|\mathbf{X}_1 = \mathbf{x}_1 \;\sim Normal_{d_2}\left(-\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{x}_1, \mathbf{V}_{22}^{-1}\right) \tag{9}$$

and

$$f_{\mathbf{X}_1}(\mathbf{x}_1) \propto \exp\left\{-\frac{1}{2}\mathbf{x}_1^\top(\mathbf{V}_{11} - \mathbf{V}_{21}^\top \mathbf{V}_{22}^{-1}\mathbf{V}_{21})\mathbf{x}_1\right\}$$

giving that

$$\mathbf{X}_1 \sim Normal_{d_1}\left(0, (\mathbf{V}_{11} - \mathbf{V}_{21}^\top \mathbf{V}_{22}^{-1}\mathbf{V}_{21})^{-1}\right). \tag{10}$$

But, from equation (3), $\Sigma_{12} = -\Sigma_{11}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}$, and then from equation (2), substituting in Σ_{12},

$$\Sigma_{11}\mathbf{V}_{11} - \Sigma_{11}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21} = \mathbf{I}_d \qquad \therefore \qquad \Sigma_{11} = (\mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21})^{-1} = (\mathbf{V}_{11} - \mathbf{V}_{21}^\top \mathbf{V}_{22}^{-1}\mathbf{V}_{21})^{-1}.$$

Hence, by inspection of equation (10), we conclude that

$$\boxed{\mathbf{X}_1 \sim Normal_{d_1}\left(0, \Sigma_{11}\right),}$$

that is, we can extract the Σ_{11} block of Σ to define the marginal sigma matrix of $\mathbf{X}_1$.

Using similar arguments, we can define the conditional distribution from equation (9) more precisely. First, from equation (3), $\mathbf{V}_{12} = -\Sigma_{11}^{-1}\Sigma_{12}\mathbf{V}_{22}$, and then from equation (5), substituting in $\mathbf{V}_{12}$

$$-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\mathbf{V}_{22} + \Sigma_{22}\mathbf{V}_{22} = \mathbf{I}_{d-d} \qquad \therefore \qquad \mathbf{V}_{22}^{-1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = \Sigma_{22} - \Sigma_{12}^\top \Sigma_{11}^{-1}\Sigma_{12}.$$

Finally, from equation (3), taking transposes on both sides, we have that $\mathbf{V}_{21}\Sigma_{11} + \mathbf{V}_{22}\Sigma_{21} = 0$. Then pre-multiplying by $\mathbf{V}_{22}^{-1}$, and post-multiplying by Σ_{11}^{-1}, we have

$$\mathbf{V}_{22}^{-1}\mathbf{V}_{21} + \Sigma_{21}\Sigma_{11}^{-1} = 0 \qquad \therefore \qquad \mathbf{V}_{22}^{-1}\mathbf{V}_{21} = -\Sigma_{21}\Sigma_{11}^{-1},$$

so we have, substituting into equation (9), that

$$\boxed{\mathbf{X}_2|\mathbf{X}_1 = \mathbf{x}_1 \;\sim Normal_{d_2}\left(\Sigma_{21}\Sigma_{11}^{-1}\mathbf{x}_1, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right).}$$

Thus any marginal, and any conditional distribution of a multivariate Normal joint distribution is also multivariate normal, as the choices of $\mathbf{X}_1$ and $\mathbf{X}_2$ are arbitrary.