14.9 Consider the family of linear Gaussian networks, as defined on page 520.
a. In a two-variable network, let X_1 be the parent of X_2, let X_1 have a Gaussian prior, and let $P(X_2 \mid X_1)$ be a linear Gaussian distribution. Show that the joint distribution $P(X_1, X_2)$ is a multivariate Gaussian, and calculate its covariance matrix.
b. Prove by induction that the joint distribution for a general linear Gaussian network on $X_1, X_2, ..., X_n$ is also a multivariate Gaussian.

a. In a two-variable network, let X_1 be the parent of X_2, let X_1 have a Gaussian prior, and let $P(X_2 \mid X_1)$ be a linear Gaussian distribution. Show that the joint distribution $P(X_1, X_2)$ is a multivariate Gaussian, and calculate its covariance matrix.

I) $P(X_1) = \frac{1}{\sigma_1 \sqrt{2\pi}} \times e^{\frac{-(x_1-\mu_1)^2}{2\sigma_1{}^2}}$

μ_1 : *mean of* X_1
σ_1 : *standard deviation of* X_1

II) $P(X_2 \mid X_1) = \frac{1}{\sigma_{x2|x1} \sqrt{2\pi}} \times e^{\frac{-(x_2-\mu_{x2|x1})^2}{2\sigma_{x2|x1}{}^2}}$

$\mu_{x2|x1}$: *mean of* X_2 *given* X_1
$\sigma_{x2|x1}$: *standard deviation of* X_2 *given* X_1

III) Mean of X_2 is a linear combination of means of Gaussian parents. X_2 presents only one parent X_1.
Then $\mu_{x_2|x_1} = ax_1 + b$. And more precisely it is:

$\mu_{x2|x1} = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1)$

μ_1 : *mean of* X_1
σ_1 : *standard deviation of* X_1
μ_2 : *mean of* X_2
σ_2 : *standard deviation of* X_2
ρ : *Rho represents correlation between* X_1 *and* X_2 $= Corr(X_1, X_2)$

A nice proof can be found in the next link:
https://newonlinecourses.science.psu.edu/stat414/node/113/

IV) TODO: prove that $\sigma_{x2|x1}$ is constant for all x1. Then:
$\sigma_{x2|x1} = \sigma_2\sqrt{(1-\rho^2)}$

Proof in:
https://newonlinecourses.science.psu.edu/stat414/node/118/

V) And now we can rewrite equation (II) using (III) and (IV):

$$P(X_2 \mid X_1) = \frac{1}{\sigma_2\sqrt{(1-\rho^2)}\sqrt{2\pi}} \times e^{\frac{-(x_2-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1)))^2}{2\sigma_2^2(1-\rho^2)}}$$

μ_1 : *mean of* X_1
σ_1 : *standard deviation of* X_1
μ_2 : *mean of* X_2
σ_2 : *standard deviation of* X_2
ρ : *Rho represents correlation between* X_1 *and* $X_2 = Corr(X_1, X_2)$

VI) Using the product rule we know that:

$$P(X_1, X_2) = P(X_2, X_1) = P(X_2 | X_1) \times P(X_1) = P(X_1) \times P(X_2 | X_1)$$

VII) We can write $P(X_1, X_2)$ as the product of equations (I) and (VII).

$$P(X_1, X_2) = \frac{1}{\sigma_1\sqrt{2\pi}} \times e^{\frac{-(x_1-\mu_1)^2}{2\sigma_1^2}} \times \frac{1}{\sigma_2\sqrt{(1-\rho^2)2\pi}} \times e^{\frac{-(x_2-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1)))^2}{2\sigma_2^2(1-\rho^2)}}$$

VIII) Resolve denominator

$$P(X_1, X_2) = \frac{1}{2\pi\times\sigma_1\times\sigma_2\times\sqrt{(1-\rho^2)}} \times e^{\frac{-(x_1-\mu_1)^2}{2\sigma_1^2}} \times e^{\frac{-(x_2-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1)))^2}{2\sigma_2^2(1-\rho^2)}}$$

IX) Resolve numerator

$$P(X_1, X_2) = \frac{1}{\sigma_1\sigma_2 2\pi} \times e^{\frac{-(x_1-\mu_1)^2}{2\sigma_1^2} + \frac{-(x_2-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1)))^2}{2\sigma_2^2(1-\rho^2)}}$$

X) Resolve exponent of e

$$= \frac{-\sigma_2^2(1-\rho^2)(x_1-\mu_1)^2}{2\sigma_1^2\sigma_2^2(1-\rho^2)} + \frac{-\sigma_1^2\,(x_2-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1)))^2}{2\sigma_1^2\sigma_2^2(1-\rho^2)}$$

$$= -\frac{1}{2(1-\rho^2)}\left[\frac{\sigma_2^2(1-\rho^2)(x_1-\mu_1)^2}{\sigma_1^2\sigma_2^2} + \frac{\sigma_1^2\,(x_2-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1)))^2}{\sigma_1^2\sigma_2^2}\right]$$

$$= -\frac{1}{2(1-\rho^2)}\left[\frac{\sigma_2^2(1-\rho^2)(x_1-\mu_1)^2 + \sigma_1^2\,((x_2-\mu_2)-\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1)))^2}{\sigma_1^2\sigma_2^2}\right]$$

$$= -\frac{1}{2(1-\rho^2)}\left[\frac{\sigma_2^2(1-\rho^2)(x_1-\mu_1)^2 + \sigma_1^2\,((x_2-\mu_2)^2 - 2\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1)(x_2-\mu_2) + (\rho\frac{\sigma_2}{\sigma_1})^2(x_1-\mu_1)^2))}{\sigma_1^2\sigma_2^2}\right]$$

$$= -\frac{1}{2(1-\rho^2)}\left[\frac{\sigma_2^2(x_1-\mu_1)^2[1-\rho^2+\rho^2] + \sigma_1^2\,((x_2-\mu_2)^2 - 2\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1)(x_2-\mu_2))}{\sigma_1^2\sigma_2^2}\right]$$

$$= -\frac{1}{2(1-\rho^2)}\left[\frac{\sigma_2^{2}(x_1-\mu_1)^2+\sigma_1^{2}(x_2-\mu_2)^2-2\rho\sigma_1\sigma_2(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1^{2}\sigma_2^{2}}\right]$$
$$= -\frac{1}{2(1-\rho^2)}\left[\frac{\sigma_2^{2}(x_1-\mu_1)^2}{\sigma_1^{2}\sigma_2^{2}}+\frac{\sigma_1^{2}(x_2-\mu_2)^2}{\sigma_1^{2}\sigma_2^{2}}-\frac{2\rho\sigma_1\sigma_2(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1^{2}\sigma_2^{2}}\right]$$
$$= -\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^{2}}+\frac{(x_2-\mu_2)^2}{\sigma_2^{2}}-\frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right]$$

XI) And now the product rule equation obtained is:

$$P(X_1, X_2) = \frac{1}{\sigma_1\sigma_2 2\pi}\times e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^{2}}+\frac{(x_2-\mu_2)^2}{\sigma_2^{2}}-\frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right]}$$

And we can see that this is exactly the Bivariate Case of a Multivariate Gaussian Distribution:
https://en.wikipedia.org/wiki/Multivariate_normal_distribution#Bivariate_case

Q.E.D. item a) 1/2

XII) Covariance in 2-dimensional case by definition:

$$\Sigma_{i,j} = E[(X_i-\mu_i)\times(X_j-\mu_j)] = Cov[X_i, X_j]$$
$$\Sigma\ = E[(X-\mu)\times(X-\mu)^T] = [Cov[X_i, X_j];\ 1\le i,\ j\le 2]$$

$\Sigma_{1,1} = E[(x_1-\mu_1)\times(x_1-\mu_1)] = \sigma_1^{2}$ by definition
$\Sigma_{1,2} = E[(x_1-\mu_1)\times(x_2-\mu_2)] = a = \rho\sigma_1\sigma_2$
$\Sigma_{2,1} = E[(x_2-\mu_2)\times(x_1-\mu_1)] = a = \rho\sigma_1\sigma_2$
$\Sigma_{2,2} = E[(x_2-\mu_2)\times(x_2-\mu_2)] = \sigma_2^{2}$ by definition

In the next link can be verified that $a = \rho\sigma_1\sigma_2$
https://en.wikipedia.org/wiki/Pearson_correlation_coefficient#For_a_population

XII) So, the covariance matrix is:

$$\Sigma = \left(\frac{\sigma_1^{2}}{\rho\sigma_1\sigma_2}\ \ \frac{\rho\sigma_1\sigma_2}{\sigma_2^{2}}\right)$$

Q.E.D. item a) 2/2

We can verify that if we start from the Multivariate Gaussian Distribution formula using the covariance found then we will arrive to the Bivariate Gaussian Case.

Multivariate Gaussian Distribution or Multivariate Normal Distribution

Multivariate Normal Distribution of a k-dimensional random vector $X = (X_1, X_2, ..., X_k)^T$:

$$X \sim N(\mu, \Sigma)$$

Mean vector:

$$\mu = E[X] = [E[X_1], E[X_2], ..., E[X_n]]^T$$

k x k covariance matrix:

$$\Sigma_{i,j} = E[(X_i - \mu_i) \times (X_j - \mu_j)] = Cov[X_i, X_j]$$
$$\Sigma = E[(X - \mu) \times (X - \mu)^T] = [Cov[X_i, X_j];\ 1 \le i,\ j \le k]$$

Density function

$$f(x = (x_1, x_2, ..., x_k)) = \frac{e^{-\frac{1}{2} \times (x-\mu)^T \times \Sigma^{-1} \times (x-\mu)}}{\sqrt{(2\pi)^k \times |\Sigma|}}$$

We need to calculate the inverse of the covariance matrix:

$$\Sigma = \begin{pmatrix} \sigma_1{}^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2{}^2 \end{pmatrix}$$
$$|\Sigma| = \sigma_1{}^2 \times \sigma_2{}^2 - \rho\sigma_1\sigma_2 \times \rho\sigma_1\sigma_2$$
$$\Sigma^{-1} = \frac{1}{\sigma_1{}^2\sigma_2{}^2 - (\rho\sigma_1\sigma_2)^2} \times \begin{pmatrix} \sigma_2{}^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1{}^2 \end{pmatrix}$$

XIII) When k=2 the Multivariate Gaussian Distribution function is:

$$f(x) = \frac{e^{-\frac{1}{2} \times (x-\mu)^T \times \Sigma^{-1} \times (x-\mu)}}{2\pi \times \sqrt{\sigma_1{}^2 \times \sigma_2{}^2 - \rho\sigma_1\sigma_2 \times \rho\sigma_1\sigma_2}}$$

$$f(x) = \frac{e^{-\frac{1}{2}\times(x-\mu)^T\times\Sigma^{-1}\times(x-\mu)}}{2\pi\times\sqrt{(\sigma_1\sigma_2)^{2}(1-\rho)}}$$

$$f(x) = \frac{e^{-\frac{1}{2}\times(x-\mu)^T\times\Sigma^{-1}\times(x-\mu)}}{2\pi\times\sigma_1\sigma_2\sqrt{(1-\rho)}}$$

So we only need to check that the exponent of e in item (XIII) is the same that is present in exponent of equation at item (XI).

XIV) Exponent of e

$$-\tfrac{1}{2}\times(x-\mu)^T\times\Sigma^{-1}\times(x-\mu)$$

$$-\tfrac{1}{2}\times(x_1-u_1;x_2-u_2)^T\times\frac{1}{\sigma_1^2\sigma_2^2-(\rho\sigma_1\sigma_2)^2}\times\begin{pmatrix}\sigma_2^2 & -\rho\sigma_1\sigma_2\\ -\rho\sigma_1\sigma_2 & \sigma_1^2\end{pmatrix}\times(x_1-u_1;x_2-u_2)$$

$$-\tfrac{1}{2}\times\frac{1}{\sigma_1^2\sigma_2^2-(\rho\sigma_1\sigma_2)^2}\times(x_1-u_1;x_2-u_2)^T\times\begin{pmatrix}\sigma_2^2 & -\rho\sigma_1\sigma_2\\ -\rho\sigma_1\sigma_2 & \sigma_1^2\end{pmatrix}\times(x_1-u_1;x_2-u_2)$$

$$-\tfrac{1}{2}\times\frac{1}{\sigma_1^2\sigma_2^2-(\rho\sigma_1\sigma_2)^2}\times[(x_1-u_1)\sigma_2^2+(x_2-u_2)(-\rho\sigma_1\sigma_2);(x_1-u_1)(-\rho\sigma_1\sigma_2)+(x_2-u_2)\sigma_1^2]\times(x_1-u_1;x_2-u_2)$$

$$-\tfrac{1}{2}\times\frac{1}{\sigma_1^2\sigma_2^2-(\rho\sigma_1\sigma_2)^2}\times[(x_1-u_1)\sigma_2^2+(x_2-u_2)(-\rho\sigma_1\sigma_2)](x_1-u_1)+[(x_1-u_1)(-\rho\sigma_1\sigma_2)+(x_2-u_2)\sigma_1^2](x_2-u_2)$$

$$-\frac{1}{2\sigma_1^2\sigma_2^2-(\rho\sigma_1\sigma_2)^2}\times\{(x_1-u_1)^2\sigma_2^2+(x_1-u_1)(x_2-u_2)(-\rho\sigma_1\sigma_2)+[(x_1-u_1)(x_2-u_2)(-\rho\sigma_1\sigma_2)+(x_2-u_2)^2\sigma_1^2]\}$$

$$-\frac{1}{2\sigma_1^2\sigma_2^2-(\rho\sigma_1\sigma_2)^2}\times\{(x_1-u_1)^2\sigma_2^2+2(x_1-u_1)(x_2-u_2)(-\rho\sigma_1\sigma_2)+(x_2-u_2)^2\sigma_1^2\}$$

$$-\frac{1}{2\sigma_1^2\sigma_2^2(1-\rho^2)}\times\{(x_1-u_1)^2\sigma_2^2+2(x_1-u_1)(x_2-u_2)(-\rho\sigma_1\sigma_2)+(x_2-u_2)^2\sigma_1^2\}$$

$$-\frac{1}{2(1-\rho^2)}\times[\frac{(x_1-u_1)^2\sigma_2^2}{\sigma_1^2\sigma_2^2}+\frac{2(x_1-u_1)(x_2-u_2)(-\rho\sigma_1\sigma_2)}{\sigma_1^2\sigma_2^2}+\frac{(x_2-u_2)^2\sigma_1^2}{\sigma_1^2\sigma_2^2}]$$

$$-\frac{1}{2(1-\rho^2)}\times[\frac{(x_1-u_1)^2}{\sigma_1^2}-\frac{2\rho(x_1-u_1)(x_2-u_2)}{\sigma_1\,\sigma_2}+\frac{(x_2-u_2)^2}{\sigma_2^2}]$$

XV) And using the exponent found in (XIV) in the original Multivariate Gaussian Distribution in (XIII) we can confirm that the covariance matrix we found is correct because we'd arrive to the Bivariate Gaussian Case!

$$f(x) = \frac{e^{-\frac{1}{2(1-\rho^2)} \times [\frac{(x_1-u_1)^2}{\sigma_1{}^2} - \frac{2\rho(x_1-u_1)(x_2-u_2)}{\sigma_1\ \sigma_2} + \frac{(x_2-u_2)^2}{\sigma_2{}^2}]}}{2\pi \times \sigma_1 \sigma_2 \sqrt{(1-\rho)}}$$

b. Prove by induction that the joint distribution for a general linear Gaussian network on $X_1, X_2, ..., X_n$ is also a multivariate Gaussian.

N=1

We suppose that for N=1 this is a Gaussian Distribution
$P(X_1)$

N=2 (this is not required for induction method but just to remember)
If $P(X_2 \mid X_1)$ is a linear Gaussian Distribution
Then $P(X_1, X_2)$ is a multivariate Gaussian Distribution
Proof in item (a) of the exercise.

N=n

We suppose that for N=n this is a general linear Bayesian Gaussian network.
Then:

- All variables are continuous
- All Conditional Probability Distributions are linear Gaussians
  - This means that the mean is a linear combinations of given variables.
  - And that the variance does not depend on given variables.

$P(X_n \mid X_1, ..., X_{n-1})$ is a linear Gaussian Distribution by Linear Gaussian Network definition
$P(X_1, ..., X_n)$ is a Multivariate Gaussian Distribution because induction method

N=n+1

Let suppose that we add one more node to the network and we reorganize it making X_{n+1} to be a child node of the entire network variables.

By product rule:
$P(X_1, ..., X_{n+1}) = P(X_{n+1} | X_1, ..., , X_n) \times P(X_1, ..., , X_n)$

By linear Gaussian Bayesian Network definition:
$P(X_{n+1} | X_1, .., , X_n) = P(X_{n+1} | Parents(X_{n+1}))$ is a linear Gaussian Distribution

By case N=n we can remember that:
$P(X_1, ..., , X_n)$ is a Multivariate Gaussian Distribution

Then $P(X_1, .., X_{n+1})$ is the result of the product of a linear Gaussian Distribution and a Multivariate

Gaussian Distribution, resulting in a new Multivariate Gaussian Distribution that presents a higher dimension.

There is another proof for this in the book “Probabilistic Graphical Models” by Koller-Friedman.

References

https://en.wikipedia.org/wiki/Multivariate_normal_distribution

https://en.wikipedia.org/wiki/Generalized_linear_model

https://en.wikipedia.org/wiki/Conjugate_prior

https://en.wikipedia.org/wiki/Exponential_family#Bayesian_estimation:_conjugate_distributions

https://vk.com/doc168073_304660839?hash=39a33dd8aa6b141d8a&dl=b667454bc650f66cc0