SDS 387D: Statistical Modeling II
Exercise 1

Travis Lilley

Jan 23, 2019

Bayesian inference in simple conjugate families

(a)

The posterior is proportional to the product of the prior and the likelihood, $f(x_1, ..., x_N|w)$. To find the likelihood, we write

$$\begin{aligned}
f(x_1, ..., x_N|w) &= \prod_{i=1}^{N} f(x_i|w) && \text{i.i.d. observations} \\
&= \prod_{i=1}^{N} w^{x_i}(1-w)^{1-x_i} && \text{density of Bernoulli} \\
&= w^{\sum_{i=1}^{N} x_i}(1-w)^{N-\sum_{i=1}^{N} x_i}. && \text{simplify product}
\end{aligned}$$

We then obtain the posterior up to a constant of proportionality.

$$\begin{aligned}
p(w|x_1, ..., x_N) &\propto f(x_1, ..., x_n|w)p(w) && \text{likelihood times prior} \\
&= \left[w^{\sum_{i=1}^{N} x_i}(1-w)^{N-\sum_{i=1}^{N} x_i}\right]\left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} w^{a-1}(1-w)^{b-1}\right] && \text{substitute} \\
&\propto w^{a+(\sum_{i=1}^{N} x_i)-1}(1-w)^{b+N-(\sum_{i=1}^{N} x_i)-1}. && \text{simplify}
\end{aligned}$$

But if we view the last expression as a function of w, then we recognize that it is the kernel of a beta density with parameters $a + \sum_{i=1}^{N} x_i$ and $b + N - \sum_{i=1}^{N} x_i$. Thus, the posterior $p(w|x_1, ..., x_N)$ is the density of a Beta$\left(a + \sum_{i=1}^{N} x_i, b + N - \sum_{i=1}^{n} x_i\right)$ distribution.

(b)

We assume that X_1 and X_2 are independent, so that their joint distribution is the product of their marginal distributions. That is,

$$\begin{aligned}
f_{X_1,X_2}(x_1,x_2) &= f_{X_1}(x_1) f_{X_2}(x_2) && \text{independence} \\
&= \left[\frac{1}{\Gamma(a_1)} x_1^{a_1-1} \exp(-x_1)\right]\left[\frac{1}{\Gamma(a_2)} x_2^{a_2-1} \exp(-x_2)\right] && \text{gamma densities} \\
&= \frac{1}{\Gamma(a_1)\Gamma(a_2)} x_1^{a_1-1} x_2^{a_2-1} \exp\{-(x_1+x_2)\}. && \text{simplify}
\end{aligned}$$

Since X_1 and X_2 are independent, the support of their joint distribution is $\mathcal{A} = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 > 0, x_2 > 0\}$.

We now consider the transformation defined by $Y_1 = \frac{X_1}{X_1+X_2}$ and $Y_2 = X_1 + X_2$. We first need to determine the image of this transformation. To do this, we solve for X_1 and X_2 in terms of Y_1 and Y_2. Since $Y_2 = X_1 + X_2$, we see that $Y_1 = \frac{X_1}{Y_2}$, so that $X_1 = Y_1Y_2$. Substituting Y_1Y_2 in for X_1 into the equation $Y_2 = X_1 + X_2$ then shows that $Y_2 = Y_1Y_2 + X_2$, so that $X_2 = Y_2 - Y_2Y_1$.

Now, since $X_2 > 0$ and $X_2 = Y_2 - Y_2Y_1$, we see that $Y_2 - Y_2Y_1 > 0$. But this implies that $Y_2 > Y_2Y_1 > 0$, which in turn implies that $0 < Y_1 < 1$. Moreover, since Y_2 is the sum of two positive real numbers, it may also equal any positive real number. Hence, the support of the joint distribution for Y_1 and Y_2 will be $\mathcal{B} = \{(y_1, y_2) \in \mathbb{R}^2 : 0 < y_1 < 1, y_2 > 0\}$.

We also need to confirm that the transformation defined by Y_1 and Y_2 is one-to-one. Suppose the points (x_1, x_2) and (x_1', x_2') in $\mathcal{A}$ are mapped to the same point in $\mathcal{B}$. It suffices to show that this implies that (x_1, x_2) and (x_1', x_2') must actually be the same point in. In other words, suppose that $\left(\frac{x_1}{x_1+x_2}, x_1 + x_2\right)$ is the same as the point $\left(\frac{x_1'}{x_1'+x_2'}, x_1'+x_2'\right)$. This implies that $\frac{x_1}{x_1+x_2} = \frac{x_1'}{x_1'+x_2'}$, and that $x_1+x_2 = x_1'+x_2'$. Hence, $\frac{x_1}{x_1+x_2} = \frac{x_1'}{x_1+x_2}$, so that $x_1 = x_1'$. And if $x_1 = x_1'$, then $x_1 + x_2 = x_1' + x_2'$ implies that $x_2 = x_2'$. Thus, (x_1, x_2) and (x_1', x_2') are actually the same point, and the transformation is one-to-one.

The final piece we need before using the formula for bivariate transformations is the Jacobian of the transformation, J. Since $X_1 = Y_1Y_2$, and $X_2 = Y_2 - Y_2Y_1$, we have

$$\begin{aligned}
J &= \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} && \text{definition of Jacobian} \\
&= \begin{vmatrix} y_2 & y_1 \\ -y_2 & 1-y_1 \end{vmatrix} && \text{partial differentiation} \\
&= y_2(1-y_1) - (y_1)(-y_2) && \text{definition of determinant} \\
&= y_2.
\end{aligned}$$

We are now ready to find the joint density of Y_1 and Y_2.

$$\begin{aligned}
f_{Y_1,Y_2}(y_1,y_2) &= f_{X_1,X_2}(x_1,x_2)|J| && \text{bivariate transformation formula} \\
&= f_{X_1,X_2}(y_1y_2, y_2 - y_2y_1)|y_2| && \text{inverse transformation and Jacobian} \\
&= \frac{y_2}{\Gamma(a_1)\Gamma(a_2)} (y_1y_2)^{a_1-1}(y_2-y_2y_1)^{a_2-1} \exp\{-(y_1y_2 + y_2 - y_2y_1)\} && \text{substitute and simplify} \\
&= \frac{1}{\Gamma(a_1)\Gamma(a_2)} y_2 y_1^{a_1-1} y_2^{a_1-1} y_2^{a_2-1} (1-y_1)^{a_2-1} \exp(-y_2) \\
&= \frac{1}{\Gamma(a_1)\Gamma(a_2)} y_1^{a_1-1}(1-y_1)^{a_2-1} y_2^{a_1+a_2-1} \exp(-y_2) \\
&= \frac{\Gamma(a_1+a_2)}{\Gamma(a_1)\Gamma(a_2)} \frac{1}{\Gamma(a_1+a_2)} y_1^{a_1-1}(1-y_1)^{a_2-1} y_2^{a_1+a_2-1} \exp(-y_2) \\
&= \underbrace{\left[\frac{\Gamma(a_1+a_2)}{\Gamma(a_1)\Gamma(a_2)} y_1^{a_1-1}(1-y_1)^{a_2-1}\right]}_{f(y_1)} \underbrace{\left[\frac{1}{\Gamma(a_1+a_2)} y_2^{a_1+a_2-1} \exp(-y_2)\right]}_{g(y_2)},
\end{aligned}$$

which has support on $\mathcal{B}$, defined earlier. But notice that we have written the joint density of Y_1 and Y_2 as the product of two functions, f and g, the first of which depends only on y_1, and the second which depends only on y_2. This means that y_1 and y_2 must be independent. Moreover, the joint density must equal the product of the two marginal densities of Y_1 and Y_2. We observe that since $0 < y_1 < 1$, and the function $f(y_1)$ is the same as the density of a Beta(a_1, a_2) distribution. Likewise, $y_2 > 0$, and $g(y_2)$ is the density of a Gamma$(a_1 + a_2, 1)$ distribution. Therefore, we conclude that, marginally,

$$Y_1 \sim \text{Beta}(a_1, a_2), \text{ and}$$
$$Y_2 \sim \text{Gamma}(a_1 + a_2, 1).$$

Suppose we wish to generate a random sample from a Beta(a_1, a_2) distribution using only random samples from gamma distributions. Then the above results suggest the following procedure:

1. Sample some X_1 from a Gamma$(a_1, 1)$ distribution.
2. Sample some X_2 from a Gamma$(a_2, 1)$ distribution that is independent of the first gamma.
3. Compute $Y^{(1)} = \frac{X_1}{X_1+X_2}$.
4. Repeat steps 1 through 3 above B times to generate $Y^{(1)}, Y^{(2)}, ..., Y^{(B)}$.

These $Y^{(1)}, Y^{(2)}, ..., Y^{(B)}$ will then constitute an i.i.d. random sample of size B from a Beta(a_1, a_2) distribution.

(c)

As in part (a), we use the fact that the posterior is proportional to the product of the likelihood and the prior of the parameter(s) of interest. Now, since the prior on θ is $N(m, v)$, we have $p(\theta) = \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{1}{2v}(\theta - m)^2\right)$. Next, we obtain the likelihood as follows:

$$
\begin{aligned}
f(x_1, ..., x_N|\theta) &= \prod_{i=1}^{N} f(x_i|\theta) && \text{i.i.d. observations} \\
&= \prod_{i=1}^{N} \left[\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2\sigma^2}(x_i - \theta)^2 \right\} \right] && \text{normal density} \\
&= \sigma^{-N}(2\pi)^{-\frac{N}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{N} (x_i - \theta)^2 \right\}. && \text{expand product and simplify}
\end{aligned}
$$

The posterior is then

$$
\begin{aligned}
p(\theta|x_1, ..., x_N) &\propto f(x_1, ..., x_N|\theta)p(\theta) && \text{Bayes' theorem} \\
&= \left[\sigma^{-N}(2\pi)^{-\frac{N}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{N}(x_i - \theta)^2 \right\} \right] \left[\frac{1}{\sqrt{2\pi v}} \exp\left\{ -\frac{1}{2v}(\theta - m)^2 \right\} \right] && \text{substitute} \\
&\propto \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{N}(x_i - \theta)^2 - \frac{1}{2v}(\theta - m)^2 \right\} && \text{drop front constants} \\
&= \exp\left\{ -\frac{1}{2}\left[\frac{1}{\sigma^2} \sum_{i=1}^{N}(x_i - \theta)^2 + \frac{1}{v}(\theta - m)^2 \right] \right\} \\
&= \exp\left\{ -\frac{1}{2}\left[\frac{1}{\sigma^2} \sum_{i=1}^{N} x_i^2 - \frac{2\theta}{\sigma^2} \sum_{i=1}^{N} x_i + \frac{1}{\sigma^2} \sum_{i=1}^{N} \theta^2 + \frac{\theta^2}{v} - \frac{2\theta m}{v} + \frac{m^2}{v} \right] \right\} && \text{expand} \\
&\propto \exp\left\{ -\frac{1}{2}\left[-\frac{2\theta}{\sigma^2} \sum_{i=1}^{N} x_i + \frac{N\theta^2}{\sigma^2} + \frac{\theta^2}{v} - \frac{2\theta m}{v} \right] \right\} && \text{drop non-}\theta\text{ terms} \\
&= \exp\left\{ -\frac{1}{2}\left[\left(\frac{N}{\sigma^2} + \frac{1}{v}\right)\theta^2 - 2\left(\frac{m}{v} + \frac{1}{\sigma^2} \sum_{i=1}^{N} x_i\right)\theta \right] \right\} && \text{group }\theta\text{ terms} \\
&= \exp\left\{ -\frac{1}{2}\left(\frac{N}{\sigma^2} + \frac{1}{v}\right)\left[\theta^2 - 2\left(\frac{N}{\sigma^2} + \frac{1}{v}\right)^{-1}\left(\frac{m}{v} + \frac{1}{\sigma^2} \sum_{i=1}^{N} x_i\right)\theta \right] \right\} && \text{complete square} \\
&\propto \exp\left\{ -\frac{1}{2}\left(\frac{N}{\sigma^2} + \frac{1}{v}\right)\left[\theta - \left(\frac{N}{\sigma^2} + \frac{1}{v}\right)^{-1}\left(\frac{m}{v} + \frac{1}{\sigma^2} \sum_{i=1}^{N} x_i\right) \right]^2 \right\} && \text{complete square}
\end{aligned}
$$

But if we view this last expression as a function of θ, we see that it is proportional to the density of a normal distribution whose mean is

$$
\begin{aligned}
m^* &= \left(\frac{N}{\sigma^2} + \frac{1}{v}\right)^{-1}\left(\frac{m}{v} + \frac{1}{\sigma^2} \sum_{i=1}^{N} x_i\right) \\
&= \frac{\sigma^2 m + v\sum_{i=1}^{N} x_i}{\sigma^2 + nv},
\end{aligned}
$$

and whose variance is

$$
\begin{aligned}
v^* &= \left(\frac{N}{\sigma^2} + \frac{1}{v}\right)^{-1} \\
&= \frac{\sigma^2 v}{\sigma^2 + nv}.
\end{aligned}
$$

Therefore, we conclude that the posterior of θ, given the data, is a Normal$\left(\frac{\sigma^2 m + v\sum_{i=1}^{N} x_i}{\sigma^2 + nv}, \frac{\sigma^2 v}{\sigma^2 + nv}\right)$ distribution.

(d)

As in parts (a) and (c), we can write the posterior, up to a constant of proportionality, as the product of the likelihood of the data and the prior of the parameter(s). The prior for ω is $p(\omega) = \frac{b^a}{\Gamma(a)}\omega^{a-1}\exp(-b\omega)$. The likelihood is

$$\begin{aligned}
f(x_1,...,x_N|\omega) &= \prod_{i=1}^{N} f(x_i|\omega) && \text{i.i.d.} \\
&= \prod_{i=1}^{N}\left[\left(\frac{\omega}{2\pi}\right)^{1/2}\exp\left\{-\frac{\omega}{2}(x_i-\theta)^2\right\}\right] && \text{gamma density} \\
&= \omega^{\frac{N}{2}}(2\pi)^{\frac{N}{2}}\exp\left\{-\frac{\omega}{2}\sum_{i=1}^{N}(x_i-\theta)^2\right\}. && \text{expand product}
\end{aligned}$$

We then can compute the posterior.

$$\begin{aligned}
p(\omega|x_1,...,x_N) &\propto f(x_1,...,x_N|\omega)p(\omega) && \text{Bayes'} \\
&= \left[\omega^{\frac{N}{2}}(2\pi)^{\frac{N}{2}}\exp\left\{-\frac{\omega}{2}\sum_{i=1}^{N}(x_i-\theta)^2\right\}\right]\left[\frac{b^a}{\Gamma(a)}\omega^{a-1}\exp(-b\omega)\right] \\
&\propto \omega^{\frac{N}{2}}\omega^{a-1}\exp\left\{-\frac{\omega}{2}\sum_{i=1}^{N}(x_i-\theta)^2\right\}\exp(-b\omega) && \text{drop initial constants} \\
&= \omega^{\frac{N}{2}+a-1}\exp\left\{-\omega\left[b+\frac{1}{2}\sum_{i=1}^{N}(x_i-\theta)^2\right]\right\}. && \text{simplify}
\end{aligned}$$

Viewing this last expression as a function of ω, we recognize it as the density of a gamma distribution whose parameters are

$$\begin{aligned}
a^* &= a+\frac{N}{2}, \text{ and} \\
b^* &= b+\frac{1}{2}\sum_{i=1}^{N}(x_i-\theta)^2.
\end{aligned}$$

Thus we see that the posterior of ω, given the data, is a Gamma($a+\frac{N}{2}, b+\frac{1}{2}\sum_{i=1}^{N}(x_i-\theta)^2$) distribution. Moreover, since $\sigma^2=\frac{1}{\omega}$, we can say that the posterior of σ^2, given the data, is IG($a+\frac{N}{2}, b+\frac{1}{2}\sum_{i=1}^{N}(x_i-\theta)^2$).

(e)

As before, we will determine the posterior by first computing the product of the prior and the likelihood. The prior for θ is $p(\theta) = \frac{1}{\sqrt{2\pi v}} \exp\left\{ -\frac{1}{2v}(\theta - m)^2 \right\}$. The likelihood is

$$\begin{aligned}
f(x_1, ..., x_N|\theta) &= \prod_{i=1}^N f(x_i|\theta) && \text{i.i.d.} \\
&= \prod_{i=1}^N \left[\frac{1}{\sigma_i\sqrt{2\pi}} \exp\left\{ -\frac{1}{2\sigma_i^2}(x_i - \theta)^2 \right\} \right] && \text{normal density} \\
&= \left(\prod_{i=1}^N \sigma_i \right) (2\pi)^{\frac{N}{2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^N \left(\frac{x_i - \theta}{\sigma_i} \right)^2 \right\}. && \text{expand product}
\end{aligned}$$

The posterior is then

$$\begin{aligned}
p(\theta|x_1, ..., x_N) &\propto f(x_1, ..., x_N|\theta)p(\theta) && \text{Bayes'} \\
&= \left[\left(\prod_{i=1}^N \sigma_i \right) (2\pi)^{\frac{N}{2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^N \left(\frac{x_i - \theta}{\sigma_i} \right)^2 \right\} \right] \left[\frac{1}{\sqrt{2\pi v}} \exp\left\{ -\frac{1}{2v}(\theta - m)^2 \right\} \right] \\
&\propto \exp\left\{ -\frac{1}{2} \sum_{i=1}^N \left(\frac{x_i - \theta}{\sigma_i} \right)^2 \right\} \exp\left\{ -\frac{1}{2v}(\theta - m)^2 \right\} && \text{drop constants} \\
&= \exp\left\{ -\frac{1}{2} \sum_{i=1}^N \frac{x_i^2}{\sigma_i^2} - \frac{1}{2} \sum_{i=1}^N \left(-\frac{2x_i\theta}{\sigma_i^2} \right) - \frac{1}{2} \sum_{i=1}^N \frac{\theta^2}{\sigma_i^2} - \frac{1}{2} \left(\frac{\theta^2}{v} \right) - \frac{1}{2} \left(-\frac{2\theta m}{v} \right) - \frac{1}{2} \left(\frac{m^2}{v} \right) \right\} \\
&\propto \exp\left\{ -\frac{1}{2} \sum_{i=1}^N \left(-\frac{2x_i\theta}{\sigma_i^2} \right) - \frac{1}{2} \sum_{i=1}^N \frac{\theta^2}{\sigma_i^2} - \frac{1}{2} \left(\frac{\theta^2}{v} \right) - \frac{1}{2} \left(-\frac{2\theta m}{v} \right) \right\} && \text{drop non-}\theta \\
&= \exp\left\{ -\frac{1}{2} \left[\left(\frac{1}{v} + \sum_{i=1}^N \frac{1}{\sigma_i^2} \right) \theta^2 - 2 \left(\frac{m}{v} + \sum_{i=1}^N \frac{x_i}{\sigma_i^2} \right) \theta \right] \right\} && \text{group } \theta \\
&= \exp\left\{ -\frac{1}{2} \left(\frac{1}{v} + \sum_{i=1}^N \frac{1}{\sigma_i^2} \right) \left[\theta^2 - 2 \left(\frac{1}{v} + \sum_{i=1}^N \frac{1}{\sigma_i^2} \right)^{-1} \left(\frac{m}{v} + \sum_{i=1}^N \frac{x_i}{\sigma_i^2} \right) \theta \right] \right\} && \text{complete square} \\
&\propto \exp\left\{ -\frac{1}{2} \left(\frac{1}{v} + \sum_{i=1}^N \frac{1}{\sigma_i^2} \right) \left[\theta - \left(\frac{1}{v} + \sum_{i=1}^N \frac{1}{\sigma_i^2} \right)^{-1} \left(\frac{m}{v} + \sum_{i=1}^N \frac{x_i}{\sigma_i^2} \right) \right]^2 \right\}. && \text{complete square}
\end{aligned}$$

As a function of theta, this last expression is proportional to the density of a normal distribution with mean and variance

$$\begin{aligned}
m^* &= \left(\frac{1}{v} + \sum_{i=1}^N \frac{1}{\sigma_i^2} \right)^{-1} \left(\frac{m}{v} + \sum_{i=1}^N \frac{x_i}{\sigma_i^2} \right) \\
&= \frac{m + v \sum_{i=1}^N x_i \sigma_i^{-2}}{1 + v \sum_{i=1}^N \sigma_i^{-2}}, \text{ and}
\end{aligned}$$

$$\begin{aligned}
v^* &= \left(\frac{1}{v} + \sum_{i=1}^N \frac{1}{\sigma_i^2} \right)^{-1} \\
&= \frac{v}{1 + v \sum_{i=1}^N \sigma_i^{-2}}.
\end{aligned}$$

Therefore, we conclude that the posterior of θ, given the data is a Normal$\left(\frac{m+v\sum_{i=1}^N x_i\sigma_i^{-2}}{1+v\sum_{i=1}^N \sigma_i^{-2}}, \frac{v}{1+v\sum_{i=1}^N \sigma_i^{-2}} \right)$ distribution.

(f)

We can first write the joint distribution of X and Ω, using definition of the conditional distribution of $x|\omega$, which states that

$$f_{X|\Omega}(x|\omega) = \frac{f_{X,\Omega}(x,\omega)}{f_\Omega(\omega)}.$$

Rearranging this last equation gives us

$$\begin{aligned} f_{X,\Omega}(x,\omega) &= f_{X|\Omega}(x|\omega) f_\Omega(\omega) \\ &= \left[\frac{\sqrt{\omega}}{\sqrt{2\pi}} \exp\left\{-\frac{\omega}{2}(x-m)^2\right\}\right]\left[\frac{(\frac{b}{2})^{\frac{a}{2}}}{\Gamma(\frac{a}{2})}\omega^{\frac{a}{2}-1}\exp\left(-\omega\frac{b}{2}\right)\right]. \end{aligned}$$

Now that we have the joint distribution of X and Ω, we can integrate out Ω to obtain the marginal distribution of X. That is,

$$\begin{aligned} f_X(x) &= \int_0^\infty f_{X,\Omega}(x,\omega) d\omega \\ &= \int_0^\infty \left[\frac{\sqrt{\omega}}{\sqrt{2\pi}} \exp\left\{-\frac{\omega}{2}(x-m)^2\right\} \frac{(\frac{b}{2})^{\frac{a}{2}}}{\Gamma(\frac{a}{2})}\omega^{\frac{a}{2}-1}\exp\left(-\omega\frac{b}{2}\right)\right] d\omega \\ &= \frac{(\frac{b}{2})^{\frac{a}{2}}}{\sqrt{2\pi}\Gamma(\frac{a}{2})} \int_0^\infty \left[\omega^{\frac{a}{2}+\frac{1}{2}-1} \exp\left\{-\omega\left(\frac{b}{2}+\frac{(x-m)^2}{2}\right)\right\}\right] d\omega \\ &= \frac{(\frac{b}{2})^{\frac{a}{2}}}{\sqrt{2\pi}\Gamma(\frac{a}{2})} \frac{\Gamma(\frac{a}{2}+\frac{1}{2})}{(\frac{b}{2}+\frac{(x-m)^2}{2})^{\frac{a}{2}+\frac{1}{2}}} \int_0^\infty \underbrace{\left[\frac{(\frac{b}{2}+\frac{(x-m)^2}{2})^{\frac{a}{2}+\frac{1}{2}}}{\Gamma(\frac{a}{2}+\frac{1}{2})}\omega^{\frac{a}{2}+\frac{1}{2}-1} \exp\left\{-\omega\left(\frac{b}{2}+\frac{(x-m)^2}{2}\right)\right\}\right]}_{\text{Gamma}\left(\frac{a}{2}+\frac{1}{2},\, \frac{b}{2}+\frac{(x-m)^2}{2}\right) \text{ density}} d\omega \\ &= \frac{(\frac{b}{2})^{\frac{a}{2}}}{\sqrt{2\pi}\Gamma(\frac{a}{2})} \frac{\Gamma(\frac{a}{2}+\frac{1}{2})}{(\frac{b}{2}+\frac{(x-m)^2}{2})^{\frac{a}{2}+\frac{1}{2}}}. \end{aligned}$$

This last expression represents the marginal distribution of X. We will show by some algebraic manipulations that this density is that of a t distribution with a shape and scale parameter. In general, such a t distribution has a pdf given by the form

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sigma\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left[1+\left(\frac{x-\mu}{\sigma\sqrt{\nu}}\right)^2\right]^{-\frac{\nu+1}{2}},$$

for all $x \in \mathbb{R}$, where $\mu \in \mathbb{R}$ is called the location parameter, $\sigma > 0$ is the scale parameter, and $\nu > 0$ is the degrees of freedom.

We rearrange the marginal density of X as follows, highlighting changes in green.

$$\begin{aligned}
f_X(x) &= \frac{(\frac{b}{2})^{\frac{a}{2}}}{\sqrt{2\pi}\Gamma(\frac{a}{2})} \frac{\Gamma(\frac{a}{2}+\frac{1}{2})}{(\frac{b}{2}+\frac{(x-m)^2}{2})^{\frac{a}{2}+\frac{1}{2}}} \\
&= \frac{(\frac{b}{2})^{\frac{a}{2}}}{\sqrt{2\pi}\Gamma(\frac{a}{2})} \frac{\Gamma(\frac{a}{2}+\frac{1}{2})}{(\frac{b}{2})^{\frac{a}{2}+\frac{1}{2}}(1+\frac{(x-m)^2}{b})^{\frac{a}{2}+\frac{1}{2}}} \\
&= \frac{(\frac{b}{2})^{-\frac{1}{2}}}{\sqrt{2\pi}\Gamma(\frac{a}{2})} \frac{\Gamma(\frac{a}{2}+\frac{1}{2})}{(1+\frac{(x-m)^2}{b})^{\frac{a}{2}+\frac{1}{2}}} \\
&= \frac{(\frac{b}{2})^{-\frac{1}{2}}\Gamma(\frac{a+1}{2})}{\sqrt{2\pi}\Gamma(\frac{a}{2})} \left[1+\frac{(x-m)^2}{b}\right]^{-\frac{a+1}{2}} \\
&= \frac{(\frac{b}{2})^{-\frac{1}{2}}\Gamma(\frac{a+1}{2})}{\sqrt{2\pi}\Gamma(\frac{a}{2})} \left[1+\left(\frac{x-m}{\sqrt{b}}\right)^2\right]^{-\frac{a+1}{2}} \\
&= \frac{(\frac{b}{2})^{-\frac{1}{2}}\Gamma(\frac{a+1}{2})}{\sqrt{2\pi}\Gamma(\frac{a}{2})} \left[1+\left(\frac{x-m}{\sqrt{\frac{b}{a}}\sqrt{a}}\right)^2\right]^{-\frac{a+1}{2}} \\
&= \frac{\Gamma(\frac{a+1}{2})}{(\frac{b}{2})^{\frac{1}{2}}\sqrt{2\pi}\Gamma(\frac{a}{2})} \left[1+\left(\frac{x-m}{\sqrt{\frac{b}{a}}\sqrt{a}}\right)^2\right]^{-\frac{a+1}{2}} \\
&= \frac{\Gamma(\frac{a+1}{2})}{\sqrt{\frac{b}{a}}\sqrt{a\pi}\Gamma(\frac{a}{2})} \left[1+\left(\frac{x-m}{\sqrt{\frac{b}{a}}\sqrt{a}}\right)^2\right]^{-\frac{a+1}{2}}.
\end{aligned}$$

But by comparing this last expression to the general form of the t distribution's pdf, we see that X must follow a t distribution with location parameter m, scale parameter $\sqrt{\frac{b}{a}}$, and a degrees of freedom.

The multivariate normal distribution

(a)

(i)

$$
\begin{aligned}
Cov(\boldsymbol{x}) &= E[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T] && \\
&= E[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}^T-\boldsymbol{\mu}^T)] && \text{linearity of transposition} \\
&= E[\boldsymbol{x}\boldsymbol{x}^T-\boldsymbol{x}\boldsymbol{\mu}^T-\boldsymbol{\mu}\boldsymbol{x}^T+\boldsymbol{\mu}\boldsymbol{\mu}^T] && \text{distributive property of matrix multiplication} \\
&= E[\boldsymbol{x}\boldsymbol{x}^T]-E[\boldsymbol{x}\boldsymbol{\mu}^T]-E[\boldsymbol{\mu}\boldsymbol{x}^T]+E[\boldsymbol{\mu}\boldsymbol{\mu}^T] && \text{linearity of expectation of random vectors} \\
&= E[\boldsymbol{x}\boldsymbol{x}^T]-\boldsymbol{\mu}\boldsymbol{\mu}^T-\boldsymbol{\mu}\boldsymbol{\mu}^T+E[\boldsymbol{\mu}\boldsymbol{\mu}^T] && \text{linearity of expectation of random vectors} \\
&= E[\boldsymbol{x}\boldsymbol{x}^T]-\boldsymbol{\mu}\boldsymbol{\mu}^T-\boldsymbol{\mu}\boldsymbol{\mu}^T+\boldsymbol{\mu}\boldsymbol{\mu}^T && \text{expectation of constants} \\
&= E[\boldsymbol{x}\boldsymbol{x}^T]-\boldsymbol{\mu}\boldsymbol{\mu}^T && \text{simplify}
\end{aligned}
$$

(ii)

$$
\begin{aligned}
&Cov(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}) = E[(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})^T]-E[\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}](E[\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}])^T && \text{result from (i)} \\
&= E[(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})(\boldsymbol{x}^T\boldsymbol{A}^T+\boldsymbol{b}^T)]-E[\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}](E[\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}])^T && \text{transposition} \\
&= E[\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{A}^T+\boldsymbol{A}\boldsymbol{x}\boldsymbol{b}^T+\boldsymbol{b}\boldsymbol{x}^T\boldsymbol{A}^T+\boldsymbol{b}\boldsymbol{b}^T]-E[\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}](E[\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}])^T && \text{distribute} \\
&= E[\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{A}^T+\boldsymbol{A}\boldsymbol{x}\boldsymbol{b}^T+\boldsymbol{b}\boldsymbol{x}^T\boldsymbol{A}^T+\boldsymbol{b}\boldsymbol{b}^T]-(\boldsymbol{A}\boldsymbol{\mu}+\boldsymbol{b})(\boldsymbol{A}\boldsymbol{\mu}+\boldsymbol{b})^T && \text{linearity of expectation} \\
&= E[\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{A}^T+\boldsymbol{A}\boldsymbol{x}\boldsymbol{b}^T+\boldsymbol{b}\boldsymbol{x}^T\boldsymbol{A}^T+\boldsymbol{b}\boldsymbol{b}^T]-(\boldsymbol{A}\boldsymbol{\mu}+\boldsymbol{b})(\boldsymbol{\mu}^T\boldsymbol{A}^T+\boldsymbol{b}^T) && \text{transposition} \\
&= E[\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{A}^T+\boldsymbol{A}\boldsymbol{x}\boldsymbol{b}^T+\boldsymbol{b}\boldsymbol{x}^T\boldsymbol{A}^T+\boldsymbol{b}\boldsymbol{b}^T]-\boldsymbol{A}\boldsymbol{\mu}\boldsymbol{\mu}^T\boldsymbol{A}^T-\boldsymbol{A}\boldsymbol{\mu}\boldsymbol{b}^T-\boldsymbol{b}\boldsymbol{\mu}^T\boldsymbol{A}^T-\boldsymbol{b}\boldsymbol{b}^T && \text{distribute} \\
&= E[\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{A}^T]+E[\boldsymbol{A}\boldsymbol{x}\boldsymbol{b}^T]+E[\boldsymbol{b}\boldsymbol{x}^T\boldsymbol{A}^T]+E[\boldsymbol{b}\boldsymbol{b}^T]-\boldsymbol{A}\boldsymbol{\mu}\boldsymbol{\mu}^T\boldsymbol{A}^T-\boldsymbol{A}\boldsymbol{\mu}\boldsymbol{b}^T-\boldsymbol{b}\boldsymbol{\mu}^T\boldsymbol{A}^T-\boldsymbol{b}\boldsymbol{b}^T && \text{distribute} \\
&= \boldsymbol{A}E[\boldsymbol{x}\boldsymbol{x}^T]\boldsymbol{A}^T+\boldsymbol{A}\boldsymbol{\mu}\boldsymbol{b}^T+\boldsymbol{b}\boldsymbol{\mu}^T\boldsymbol{A}^T+\boldsymbol{b}\boldsymbol{b}^T-\boldsymbol{A}\boldsymbol{\mu}\boldsymbol{\mu}^T\boldsymbol{A}^T-\boldsymbol{A}\boldsymbol{\mu}\boldsymbol{b}^T-\boldsymbol{b}\boldsymbol{\mu}^T\boldsymbol{A}^T-\boldsymbol{b}\boldsymbol{b}^T && \text{expectation properties} \\
&= \boldsymbol{A}E[\boldsymbol{x}\boldsymbol{x}^T]\boldsymbol{A}^T-\boldsymbol{A}\boldsymbol{\mu}\boldsymbol{\mu}^T\boldsymbol{A}^T && \text{simplify} \\
&= \boldsymbol{A}(E[\boldsymbol{x}\boldsymbol{x}^T]-\boldsymbol{\mu}\boldsymbol{\mu}^T)\boldsymbol{A}^T && \text{factor} \\
&= \boldsymbol{A}Cov(\boldsymbol{x})\boldsymbol{A}^T && \text{result from (i)}
\end{aligned}
$$

(b)

The density of $\boldsymbol{z}$ is the joint distribution of its individual random variables. That is $f(\boldsymbol{z}) = f_{Z_1,\dots,Z_p}(z_1,\dots,z_p)$. But since the Z_i are independent, their joint distribution is simply the product of their marginal distributions. That is, $f_{Z_1,\dots,Z_p} = \prod_{i=1}^p f_{Z_i}(z_i)$. Moreover, since the Z_i are all identically normal with mean 0 and variance 1, we have $f_{Z_i}(z_i) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}z_i^2\}$ for all i, so that

$$
\begin{aligned}
f(\boldsymbol{z}) &= \prod_{i=1}^p f_{Z_i}(z_i) && \text{indp.}\\
&= \prod_{i=1}^p \frac{1}{\sqrt{2\pi}} \exp\Big\{-\frac{1}{2}z_i^2\Big\} && \text{identical normals}\\
&= \frac{1}{(2\pi)^{\frac{p}{2}}} \exp\Big\{-\frac{1}{2}z_1^2 - \cdots - \frac{1}{2}z_p^2\Big\} && \text{property of exp()}\\
&= \frac{1}{(2\pi)^{\frac{p}{2}}} \exp\Big\{-\frac{1}{2}(z_1^2 + \cdots z_p^2)\Big\} && \text{simplify}\\
&= \frac{1}{(2\pi)^{\frac{p}{2}}} \exp\Big\{-\frac{1}{2}\boldsymbol{z}^T\boldsymbol{z}\Big\}, && \text{inner product}
\end{aligned}
$$

where $\boldsymbol{z} \in \mathbb{R}^p$.

Next, using the definition of the mgf of a random vector, we have

$$
\begin{aligned}
m_{\boldsymbol{z}}(\boldsymbol{t}) &= E_{\boldsymbol{z}}[\exp\{\boldsymbol{t}^T\boldsymbol{z}\}] && \text{definition}\\
&= E_{\boldsymbol{z}}\Big[\exp\Big\{\sum_{i=1}^p t_i Z_i\Big\}\Big] && \text{inner product}\\
&= E_{\boldsymbol{z}}\Big[\prod_{i=1}^p \exp\Big\{t_i Z_i\Big\}\Big] && \text{property of exp()}\\
&= \prod_{i=1}^p E_{Z_i}\Big[\exp\{t_i Z_i\}\Big] && \text{expectation of ind. r.v.'s}\\
&= \prod_{i=1}^p \Big[\int_{\mathbb{R}} \Big(\exp\{t_i z_i\} \frac{1}{\sqrt{2\pi}} \exp\Big\{-\frac{1}{2}z_i^2\Big\}\Big) dz_i\Big] && \text{def. of univar expectation}\\
&= \prod_{i=1}^p \Big[\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\Big\{-\frac{1}{2}z_i^2 + t_i z_i\Big\} dz_i\Big] && \text{combine exp() terms}\\
&= \prod_{i=1}^p \Big[\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\Big\{-\frac{1}{2}(z_i^2 - 2t_i z_i + t_i^2 - t_i^2)\Big\} dz_i\Big] && \text{complete square}\\
&= \prod_{i=1}^p \Big[\exp\Big\{\frac{1}{2}t_i^2\Big\} \int_{\mathbb{R}} \underbrace{\frac{1}{\sqrt{2\pi}} \exp\Big\{-\frac{1}{2}(z_i - t_i)^2\Big\}}_{\text{density of } N(t_i,1) \text{ dist'n}} dz_i\Big] && \text{complete square, pull out const}\\
&= \prod_{i=1}^p \exp\Big\{\frac{1}{2}t_i^2\Big\} && \text{integral of density} = 1\\
&= \exp\Big\{\sum_{i=1}^p \frac{1}{2}t_i^2\Big\} && \text{property of exp()}\\
&= \exp\Big\{\frac{1}{2}\boldsymbol{t}^T\boldsymbol{t}\Big\}, && \text{inner product}
\end{aligned}
$$

for all $\boldsymbol{t} \in \mathbb{R}^p$.

(c)

From the definition beginning at the problem, we see that $Y = \boldsymbol{t}^T\boldsymbol{x}$, as a linear combination of univariate normals, must be a univariate normal distribution itself. Moreover, using our results from earlier, we can obtain the mean and variance of Y as follows:

$$\begin{aligned} E[Y] &= E[\boldsymbol{t}^T\boldsymbol{x}] \\ &= \boldsymbol{t}^T E[\boldsymbol{x}] && \text{part (a)} \\ &= \boldsymbol{t}^T\boldsymbol{\mu}. \end{aligned}$$

$$\begin{aligned} Var(Y) &= Var(\boldsymbol{t}^T\boldsymbol{x}) \\ &= \boldsymbol{t}^T Cov(\boldsymbol{x})\boldsymbol{t}^T && \text{part (a)} \\ &= \boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}. \end{aligned}$$

But from the exercise instructions, we know that the mgf of Y must be $m_Y(s) = \exp\{\mu t + \frac{s^2\sigma^2}{2}\}$, where μ and σ^2 are its mean and variance. Therefore,

$$\begin{aligned} m_Y(s) &= E[\exp\{sY\}] \\ &= \exp\left\{\mu s + \frac{s^2\sigma^2}{2}\right\} && \text{univariate normal mgf} \\ &= \exp\left\{\boldsymbol{t}^T\boldsymbol{\mu}s + \frac{s^2\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}}{2}\right\}. && \text{result from above} \end{aligned}$$

Now, suppose that we evaluate $m_Y(s)$ at $s = 1$. This would given us

$$\begin{aligned} m_Y(1) &= E[\exp\{(1)(Y)\}] \\ &= E[Y] \\ &= E[\boldsymbol{t}^T\boldsymbol{x}] \\ &= \exp\left\{\boldsymbol{t}^T\boldsymbol{\mu}(1) + \frac{(1)^2\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}}{2}\right\} \\ &= \exp\left\{\boldsymbol{t}^T\boldsymbol{\mu} + \frac{\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}}{2}\right\}. \end{aligned}$$

Thus, $E[\boldsymbol{t}^T\boldsymbol{x}] = \exp\{\boldsymbol{t}^T\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}\}$, as desired, and we conclude that this is the mgf of a multivariate normal $\boldsymbol{x}$. Furthermore, since the mgf of any distribution uniquely characterizes it, we know that if any other distribution has this same mgf, then it, too, must be multivariate normal. Hence, we can define $\boldsymbol{x}$ as being multivariate normal if and only if its mgf is $\exp\{\boldsymbol{t}^T\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}\}$.

(d)

First, we show that if $\boldsymbol{x}$ is a multivariate normal distribution, and $\boldsymbol{c}$ is a (additively conformable) vector of constants, then $\boldsymbol{x}+\boldsymbol{c}$ is also multivariate normal. To see why, we use the mgf developed in the previous problem. Suppose $\boldsymbol{x}$ has mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. From part (c), we know that its mgf must be

$$m_{\boldsymbol{x}}(\boldsymbol{t}) = \exp\left\{\boldsymbol{t}^T\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}\right\}.$$

The mgf of $\boldsymbol{x}+\boldsymbol{c}$ is then

$$\begin{aligned}
m_{\boldsymbol{x}+\boldsymbol{c}}(t) &= E[\exp\{\boldsymbol{t}^T(\boldsymbol{x}+\boldsymbol{c})\}] && \text{definition of mgf}\\
&= E[\exp\{\boldsymbol{t}^T\boldsymbol{x}\}\exp\{\boldsymbol{t}^T\boldsymbol{c}\}] && \text{property of exp()}\\
&= \exp\{\boldsymbol{t}^T\boldsymbol{c}\}E[\exp\{\boldsymbol{t}^T\boldsymbol{x}\}] && \text{linearity of E[]}\\
&= \exp\{\boldsymbol{t}^T\boldsymbol{c}\}\exp\left\{\boldsymbol{t}^T\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}\right\} && \text{mgf of multivar normal}\\
&= \exp\left\{\boldsymbol{t}^T(\boldsymbol{c}+\boldsymbol{\mu}) + \frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}\right\} && \text{property of exp().}
\end{aligned}$$

But from the results of part (c), we know that this means $\boldsymbol{x}$ must be multivariate normal with mean $\boldsymbol{\mu}+\boldsymbol{c}$ and covariance $\boldsymbol{\Sigma}$.

Now, suppose that $\boldsymbol{x} = \boldsymbol{L}\boldsymbol{z}+\boldsymbol{\mu}$, where $\boldsymbol{z}$ is standard multivariate normal, and $\boldsymbol{\mu}$ is a vector of constants. If we can show that $\boldsymbol{a}^T\boldsymbol{x}$ is univariate normal for any nonzero $\boldsymbol{a}$, then the definition at the beginning of part (c) suggests that $\boldsymbol{x}$ must be multivariate normal. We observe that $\boldsymbol{a}^T\boldsymbol{x} = \boldsymbol{a}^T[\boldsymbol{L}\boldsymbol{z}+\boldsymbol{\mu}] = \boldsymbol{a}^T\boldsymbol{L}\boldsymbol{z} + \boldsymbol{a}^T\boldsymbol{\mu}$. But since $\boldsymbol{z}$ is multivariate normal, we know that $\boldsymbol{a}^T\boldsymbol{L}\boldsymbol{z}$, as a linear combination of the random vectors in $\boldsymbol{z}$, must be univariate normal. Also, we showed earlier in this problem that adding a constant to a normal distribution results in another normal distribution. (We showed this in general for multivariate normal distributions, of which univariate normals are special case.) Therefore, $\boldsymbol{a}^T\boldsymbol{L}\boldsymbol{z}+\boldsymbol{a}^T\boldsymbol{\mu}$ is univariate normal, and we see that $\boldsymbol{x}$ must be multivariate normal.

To find the mean and variance of $\boldsymbol{x}$, we use results from part (a). Recall that the mean of the standard multivariate normal distribution is $\mathbf{0}$, and that its variance is $\boldsymbol{I}$, the identity matrix.

$$\begin{aligned}
E[\boldsymbol{x}] &= E[\boldsymbol{L}\boldsymbol{z}+\boldsymbol{\mu}]\\
&= \boldsymbol{L}E[\boldsymbol{z}]+\boldsymbol{\mu} && \text{property from (a)}\\
&= \boldsymbol{L}(\mathbf{0})+\boldsymbol{\mu} && \text{mean of } \boldsymbol{z}\\
&= \boldsymbol{\mu}.
\end{aligned}$$

$$\begin{aligned}
Var[\boldsymbol{x}] &= Var[\boldsymbol{L}\boldsymbol{z}+\boldsymbol{\mu}]\\
&= Var[\boldsymbol{L}\boldsymbol{z}] && \text{variance unchanged by addition of const}\\
&= \boldsymbol{L}Var[\boldsymbol{z}]\boldsymbol{L}^T && \text{property from (b)}\\
&= \boldsymbol{L}\boldsymbol{I}\boldsymbol{L}^T && \text{variance of } \boldsymbol{z}\\
&= \boldsymbol{L}\boldsymbol{L}^T.
\end{aligned}$$

Hence, $\boldsymbol{x}$ is multivariate normal with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{L}\boldsymbol{L}^T$.

(e)

Let $\boldsymbol{x}$ be multivariate normal with mean μ and covariance matrix $\boldsymbol{\Sigma}$. Then $\boldsymbol{\Sigma}$, like every covariance matrix, is positive definite, so that its eigen-decomposition can be written as $\boldsymbol{\Sigma} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$, where $\boldsymbol{Q}$ is an orthogonal matrix containing the eigenvectors of $\boldsymbol{\Sigma}$, and $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{\Sigma}$. Moreover, since $\boldsymbol{\Sigma}$ is positive definite, all of its eigenvalues are positive. Hence, we can write $\boldsymbol{\Lambda}$ as $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Lambda}^{\frac{1}{2}}$, where $\boldsymbol{\Lambda}^{\frac{1}{2}}$ is a diagonal matrix containing the square roots of the eigenvalues of $\boldsymbol{\Sigma}$. Hence, $\boldsymbol{\Sigma} = \boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{Q}^T$.

Now, suppose that $\boldsymbol{z}$ is the standard multivariate normal distribution. From the results of part (d), we know that the affine transformation $\boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{z} + \boldsymbol{\mu}$ must result is a multivariate normal distribution. We will show that this transformation results in $\boldsymbol{x}$, i.e., that $\boldsymbol{x} = \boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{z} + \boldsymbol{\mu}$. From our earlier results, it is sufficient to show that the mean and variance of $\boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{z} + \boldsymbol{\mu}$ are the same as that of $\boldsymbol{x}$ in order to establish that $\boldsymbol{x} = \boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{z} + \boldsymbol{\mu}$. We verify this now.

$$\begin{aligned}
E[\boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{z} + \boldsymbol{\mu}] &= \boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}E[\boldsymbol{z}] + \boldsymbol{\mu} \\
&= \boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{0} + \boldsymbol{\mu} \\
&= \boldsymbol{\mu}.
\end{aligned}$$

$$\begin{aligned}
Var[\boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{z} + \boldsymbol{\mu}] &= Var[\boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{z}] \\
&= \boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}Var[\boldsymbol{z}](\boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}})^T \\
&= \boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{I}(\boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}})^T \\
&= \boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}(\boldsymbol{\Lambda}^{\frac{1}{2}})^T\boldsymbol{Q}^T \\
&= \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T \\
&= \boldsymbol{\Sigma}.
\end{aligned}$$

Therefore, $\boldsymbol{x}$ does indeed equal $\boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{z} + \boldsymbol{\mu}$, so that we can indeed express any multivariate normal distribution $\boldsymbol{x}$ as an affine transformation of the standard multivariate normal distribution.

(f)

Again, suppose that $\boldsymbol{x}$ is multivariate normal with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. We consider the transformation $\boldsymbol{x} = \boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{z} + \boldsymbol{\mu}$, where $\boldsymbol{\Sigma} = \boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{Q}^T$ is the eigendecomposition of $\boldsymbol{\Sigma}$, and $\boldsymbol{z}$ is the standard multivariate normal distribution. We will use the formula for pdf of transformation random variables to derive the pdf of $\boldsymbol{x}$. Recall that the pdf of $\boldsymbol{z}$ is

$$f_{\boldsymbol{z}}(\boldsymbol{z}) = \frac{1}{(2\pi)^{\frac{p}{2}}} \exp\left\{ -\frac{1}{2}\boldsymbol{z}^T\boldsymbol{z} \right\},$$

for all $\boldsymbol{z} \in \mathbb{R}^p$. Next, observe that the inverse of the transformation, in which we solve for $\boldsymbol{z}$ in terms of $\boldsymbol{x}$ is given by $\boldsymbol{z} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$. From the instructions, we know that the Jacobian of this transformation is then $\boldsymbol{J} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}^{-1}$. Using the formula for transformation of random variables, we then have

$$\begin{aligned}
f_{\boldsymbol{x}}(\boldsymbol{x}) &= f_{\boldsymbol{z}}(\boldsymbol{z})|\boldsymbol{J}| \\
&= f_{\boldsymbol{z}}(\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}))|\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}^{-1}| \\
&= |\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}^{-1}|(2\pi)^{-\frac{p}{2}} \exp\left\{ -\frac{1}{2}[\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})]^T[\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})] \right\} \\
&= |\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}^{-1}|(2\pi)^{-\frac{p}{2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T(\boldsymbol{Q}^{-1})^T(\boldsymbol{\Lambda}^{-\frac{1}{2}})^T\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \right\} \\
&= |\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}^{-1}|(2\pi)^{-\frac{p}{2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T(\boldsymbol{Q}^T)^{-1}(\boldsymbol{\Lambda}^T)^{-\frac{1}{2}}\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \right\} \\
&= |\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}^{-1}|(2\pi)^{-\frac{p}{2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T(\boldsymbol{Q}^T)^{-1}\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \right\} \\
&= |\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}^{-1}|(2\pi)^{-\frac{p}{2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \right\} \\
&= |\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}^{-1}|(2\pi)^{-\frac{p}{2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \right\} \\
&= |\boldsymbol{\Sigma}^{-1}|^{\frac{1}{2}}(2\pi)^{-\frac{p}{2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \right\} \\
&= |\boldsymbol{\Sigma}|^{-\frac{1}{2}}(2\pi)^{-\frac{p}{2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \right\} \\
&= \frac{1}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \right\}.
\end{aligned}$$

Therefore, for a multivariate random vector $\boldsymbol{x}$ with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, its given is given by the last expression.

(g)

We proceed by finding the mgf of $\boldsymbol{Ax}_1 + \boldsymbol{Bx}_2$. We know from our earlier results that both $\boldsymbol{Ax}_1$ and $\boldsymbol{Bx}_2$ are both multivariate normal. Also, using other previous results, we have that the mean and covariance of $\boldsymbol{Ax}_1$ are $\boldsymbol{A\mu}_1$ and $\boldsymbol{A\Sigma}_1\boldsymbol{A}^T$, respectively. Similarly, the mean and covariance of $\boldsymbol{Bx}_2$ are $\boldsymbol{B\mu}_2$ and $\boldsymbol{B\Sigma}_2\boldsymbol{B}^T$.

$$
\begin{aligned}
m_{\boldsymbol{Ax}_1+\boldsymbol{Bx}_2}(\boldsymbol{t}) &= E[\exp\{\boldsymbol{t}^T(\boldsymbol{Ax}_1 + \boldsymbol{Bx}_2)\}] \\
&= E[\exp\{\boldsymbol{t}^T\boldsymbol{Ax}_1\}\exp\{\boldsymbol{t}^T\boldsymbol{Bx}_2\}] \\
&= E[\exp\{\boldsymbol{t}^T\boldsymbol{Ax}_1\}]E[\exp\{\boldsymbol{t}^T\boldsymbol{Bx}_2\}] \\
&= m_{\boldsymbol{Ax}_1}(\boldsymbol{t}) \cdot m_{\boldsymbol{Bx}_2}(\boldsymbol{t}) \\
&= \exp\left\{\boldsymbol{t}^T\boldsymbol{A\mu}_1 + \frac{1}{2}\boldsymbol{t}^T\boldsymbol{A\Sigma}_1\boldsymbol{A}^T\boldsymbol{t}\right\} \cdot \exp\left\{\boldsymbol{t}^T\boldsymbol{B\mu}_2 + \frac{1}{2}\boldsymbol{t}^T\boldsymbol{B\Sigma}_2\boldsymbol{B}^T\boldsymbol{t}\right\} \\
&= \exp\left\{\boldsymbol{t}^T\boldsymbol{A\mu}_1 + \frac{1}{2}\boldsymbol{t}^T\boldsymbol{A\Sigma}_1\boldsymbol{A}^T\boldsymbol{t} + \boldsymbol{t}^T\boldsymbol{B\mu}_2 + \frac{1}{2}\boldsymbol{t}^T\boldsymbol{B\Sigma}_2\boldsymbol{B}^T\boldsymbol{t}\right\} \\
&= \exp\left\{\boldsymbol{t}^T(\boldsymbol{A\mu}_1 + \boldsymbol{B\mu}_2) + \frac{1}{2}\boldsymbol{t}^T(\boldsymbol{A\Sigma}_1\boldsymbol{A}^T + \boldsymbol{B\Sigma}_2\boldsymbol{B}^T)\boldsymbol{t}\right\},
\end{aligned}
$$

which we know from our earlier results means that $\boldsymbol{Ax}_1+\boldsymbol{Bx}_2$ must be multivariate normal with mean $\boldsymbol{A\mu}_1+\boldsymbol{B\mu}_2$ and covariance $\boldsymbol{A\Sigma}_1\boldsymbol{A}^T + \boldsymbol{B\Sigma}_2\boldsymbol{B}^T$.