Solution 2 - Multivariate Distributions

1. In publicly available solution manual.

   Supplemental material & typos in the solution manual

   - **Murphy 2.4 [Convolution of two Gaussians is a Gaussian]**
     There is a typo in the solution to this question. Instead of $\sigma_2^2\sigma_2^2$, it should be $\sigma_1^2\sigma_2^2$.
     i.e., equation (9) should be:

$$\int \exp\left[-\frac{1}{2}(w_1+w_2)(x_1-\hat{x})^2\right] dx_1 = (2\pi)^{\frac{1}{2}}\left(\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right)^{\frac{1}{2}}$$

     and equation (10) should be:

$$p(y) = (2\pi)^{-1}(\sigma_1^2\sigma_2^2)^{-\frac{1}{2}}(2\pi)^{\frac{1}{2}}\left(\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right)^{\frac{1}{2}}\exp\left[-\frac{1}{2(\sigma_1^2+\sigma_2^2)}(y-\mu_1-\mu_2)^2\right]$$

Another way to get equation (8) in the solution:

$$\begin{aligned}
p(y) &= \int_{-\infty}^{\infty} p(x_1)p(y-x_1)\,dx_1 \quad \text{(by the definition of convolution)}\\
&= \int_{-\infty}^{\infty}\left[\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right)\right]\left[\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{(y-x_1-\mu_2)^2}{2\sigma_2^2}\right)\right]dx_1\\
&= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty}\exp\left[-\frac{(x_1-\mu)^2}{2\sigma_1^2}-\frac{(y-x_1-\mu_2)^2}{2\sigma_2^2}\right]dx_1
\end{aligned}$$

Let $w_1 = \frac{1}{\sigma_1^2}, w_2 = \frac{1}{\sigma_2^2}$, we have

$$\begin{aligned}
p(y) &= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}\left(w_1(x_1-\mu)^2+w_2(y-x_1-\mu_2)^2\right)\right]dx_1\\
&= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}\left[(w_1+w_2)x_1^2-2(\mu_1w_1+w_2y-\mu_2w_2)x_1+\mu_1^2w_1+\mu_2^2w_2+w_2y^2-2\mu_2w_2y\right]\right]dx_1
\end{aligned}$$

Let $\hat{x} = \frac{\mu_1w_1+w_2y-\mu_2w_2}{w_1+w_2}$, we have

$$\begin{aligned}
p(y) &= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}\left((w_1+w_2)(x_1-\hat{x})^2+\frac{w_1w_2}{w_1+w_2}(\mu_1^2+\mu_2^2-2\mu_2y-2\mu_1y+2\mu_1\mu_2)\right)\right]dx_1\\
&= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}(w_1+w_2)(x_1-\hat{x})^2-\frac{1}{2}\frac{w_1w_2}{w_1+w_2}(y-(\mu_1-\mu_2))^2\right]dx_1
\end{aligned}$$

- **Murphy 3.2[Correlation coefficient is between -1 and +1]**
  Another way to prove it:
  Let $Z = X - aY$,

$$\begin{aligned}
V[Z] &= E[Z^2] - (E[Z]^2) \\
&= E[X^2 + a^2Y^2 - 2aXY] - (\mu_X^2 + a^2\mu_Y^2 - 2a\mu_X\mu_Y) \\
&= (E[X^2] - \mu_X^2) + (a^2E[Y^2] - a^2\mu_Y^2) - (2aE[XY] - 2a\mu_X\mu_Y) \\
&= V[X] + a^2V[Y] - 2a\mathrm{Cov}[X,Y]
\end{aligned}$$

  Since $V[Z] \geq 0$, we have $V[X] + a^2V[Y] \geq 2a\mathrm{Cov}[X,Y]$, $\forall a \in \mathbb{R}$.
  When we choose $a = \frac{\sigma_X}{\sigma_Y}$,

$$\mathrm{Cov}[X,Y] \leq \sigma_X\sigma_Y \Longrightarrow \rho_{X,Y} \leq 1$$

  When we choose $a = -\frac{\sigma_X}{\sigma_Y}$,

$$\mathrm{Cov}[X,Y] \geq -\sigma_X\sigma_Y \Longrightarrow \rho_{X,Y} \geq -1$$

- **Murphy 3.4 [Linear combinations of random variables]**
  More details on part(c):

$$\begin{aligned}
E[\mathbf{x}^\top\mathbf{A}\mathbf{x}] &= E[\mathrm{tr}(\mathbf{x}^\top\mathbf{A}\mathbf{x})] \quad \text{(because } \mathbf{x}^\top\mathbf{A}\mathbf{x} \text{ is a } 1\times 1 \text{ matrix, i.e. a scalar)} \\
&= E[\mathrm{tr}(\mathbf{A}\mathbf{x}\mathbf{x}^\top)] \quad \text{(based on part(b), we can switch } \mathbf{x}^\top \text{ and } \mathbf{A}\mathbf{x}) \\
&= \mathrm{tr}(\mathbf{A}E[\mathbf{x}\mathbf{x}^\top]) \quad \text{(because } E[\cdot] \text{ and } \mathrm{tr}(\cdot) \text{ are both linear operators)} \\
&= \mathrm{tr}(\mathbf{A}(\mathbf{\Sigma} + \mathbf{m}\mathbf{m}^\top)) \quad (\because \mathrm{Cov}(\mathbf{x},\mathbf{x}) = E[\mathbf{x}\mathbf{x}^\top] - E[\mathbf{x}]E[\mathbf{x}]^\top) \\
&= \mathrm{tr}(\mathbf{A}\mathbf{\Sigma}) + \mathrm{tr}(A\mathbf{m}\mathbf{m}^\top) \quad \text{(linearity of trace)} \\
&= \mathrm{tr}(\mathbf{A}\mathbf{\Sigma}) + \mathrm{tr}(\mathbf{m}^\top\mathbf{A}\mathbf{m}) \quad \text{(based on part(b))} \\
&= \mathrm{tr}(\mathbf{A}\mathbf{\Sigma}) + \mathbf{m}^\top\mathbf{A}\mathbf{m} \quad (\because \mathbf{m}^\top\mathbf{A}\mathbf{m} \text{ is a scalar})
\end{aligned}$$

- **Murphy 3.5 [Gaussian vs *jointly* Gaussian]**
  (a) Equation (50) should be:

$$E[Y] = E[WX] = E[W]E[X] \quad (\because W, X \text{ are independent})$$

  Another way to calculate $V[Y]$:

$$\begin{aligned}
V[Y] &= E[Y^2] - E[Y]^2 \\
&= E[(WX)^2] - E[WX]^2 \\
&= E[W^2]E[X^2] - (E[W]E[X])^2 \\
&= E[W^2](V[X] + E[X]^2) - (E[W]E[X])^2 \\
&= E[W^2](1+0) - 0 \\
&= E[W^2] = 1
\end{aligned}$$

  (b) correction for typos in Equation (56) and Equation (57):

$$\begin{aligned}
E[XY] = \cdots &= \sum_{w\in\{-1,1\}} p(w)E[XY|W] \\
&= 0.5 \cdot E[X\cdot(-X)] + 0.5\cdot E[X\cdot X] \\
&= 0
\end{aligned}$$

- **Murphy 3.6 [Normalization constant for a multidimensional Gaussian]**
  Equation (68) should be:
$$\prod_i \sqrt{2\pi\lambda_i} = (\sqrt{2\pi})^d |\mathbf{\Sigma}|^{\frac{1}{2}} = (2\pi)^{d/2} |\mathbf{\Sigma}|^{\frac{1}{2}}$$

2.

$$\begin{aligned} E[E[X \mid Y]] &= \int_{-\infty}^{\infty} E[X \mid Y = y] f_Y(y)\, dy \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x f_{X|Y}(x \mid y)\, dx \right) f_Y(y)\, dy \end{aligned}$$

Rearranging terms in the double integral and reversing the order of integration, we obtain:

$$\begin{aligned} E[E[X \mid Y]] &= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{X|Y}(x \mid y) f_Y(y)\, dy\, dx \\ &= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy\, dx \quad \text{(by Bayes' rule)} \\ &= \int_{-\infty}^{\infty} x f_X(x)\, dx \quad \text{(by marginalizing out Y)} \end{aligned}$$

3.

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy = \int_0^x 2\, dy = 2x$$

The conditional PDF of Y given X is:

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \begin{cases} 1/x & 0 \le y \le x, \\ 0 & \text{otherwise.} \end{cases}$$

Given $X = x$, we see that Y is the uniform $(0, x)$ random variable.

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx = \int_y^1 2\, dx = 2(1 - y)$$

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \begin{cases} 1/(1 - y) & y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Conditioned on $Y = y$, we see that X is the uniform $(y, 1)$ random variable.